Modeling of Craniofacial Anatomy, Variation, and Growth

Thorup, Signe Strann

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Modeling of Craniofacial Anatomy, Variation, and Growth

Signe Strann Thorup

Kongens Lyngby 2013
IMM-PHD-2013-290

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

IMM-PHD: ISSN 0909-3192

Summary

The topic of this thesis is automatic analysis of craniofacial images with respect to changes due to growth and surgery, inter-subject variation and intracranial volume estimation. The methods proposed contribute to the knowledge about specific craniofacial anomalies, as well as provide a tool for detailed analyses for clinical and research purposes.

Most of the applications in this thesis rely on non-rigid image registration by the means of warping one image into the coordinate system of another image. This warping results in a deformation field that describes the anatomical correspondence between the two images. To elaborate further: a computational atlas of the average anatomy was constructed. Using non-rigid registration, image data from a subject is automatically transformed into the coordinate space of the atlas. In this process, all knowledge built into the atlas is transferred to the subject, thus creating a personalized atlas. The knowledge built into the atlas is e.g. location of anatomical regions and landmarks of importance to surgery planning and evaluation or population studies. With these correspondences, various analyses could be carried out e.g. quantification of growth, inter-subject variation etc. Besides image registration, a volumetric segmentation method using graph cuts was developed and applied for intracranial volume estimation. Graph cut is a fast method for segmentation utilizing a suitable graph.

Three different craniofacial anomalies were examined in this thesis: Cleft lip and palate, unicoronal synostosis, and Crouzon syndrome. Using the proposed methods, highly detailed variation was assessed for cleft lip and palate, correspondence between images obtained before and after lip repair was established for cleft lip and palate, the intracranial volume was estimated for infants with

unicoronal synostosis, and finally, craniofacial growth patterns were quantified for Crouzon syndrome in a mouse model.

Resumé

Denne afhandling omhandler automatisk analyse af kraniofaciale billeder mht. forandringer forårsaget af vækst og kirurgiske indgreb, variation mellem individer og estimering af det intrakranielle volumen. Metoderne bidrager til forståelsen af specifikke kraniofaciale misdannelser, og fungerer som værktøj ifm. detaljerede analyser til kliniske og forskningsbaserede formål.

De fleste anvendelser i afhandlingen beror på ikke-rigid billedregistrering, hvor et billede transformeres over i et andet billedes koordinatsystem. Denne transformering resulterer i et deformationsfelt, som beskriver den anatomiske korrespondance mellem de to billeder. For at uddybe dette nærmere blev et atlas af den gennemsnitlige anatomi konstrueret. Ikke-rigid billed-registrering blev herefter brugt til automatisk at transformere billeddata fra et individ til atlassets koordinatsystem. I denne proces blev den anatomiske viden indbygget i atlasset overført til individet, for på den måde at generere et personligt atlas. Den anatomiske viden i atlasser kan f.eks. være specifikke anatomiske deles placering eller vigtige referencepunkter til brug i f.eks. operationsplanlægning, operationsevaluering eller populationsstudier. Korrespondancerne muliggør adskillelige analyser f.eks. kvantificering af vækst, variation mellem individer osv.
Udover billedregistrering blev en volumetrisk segmenterings metode baseret på graph cuts udviklet og brugt til estimering af det intrakranielle volumen hos patienter. Graph cut er en hurtig metode til segmentering såfremt en passende graf er sat op.

I denne afhandling blev tre forskellige kraniofaciale misdannelser undersøgt: Læbe-ganespalte, enkeltsidig coronal synostose og Crouzon syndrom. Ved brug af de fremsatte metoder blev variation undersøgt i høj detaljeringsgrad for spædbørn med læbe-ganespalte, korrespondance blev etableret mellem billedet optaget hhv.

før og efter kirurgisk lukning af læbespalten hos spædbørn med læbe-ganespalte. Det intrakranielle volumen blev estimeret hos spædbørn med enkeltsidig coronal synostose. Endvidere blev kraniofaciale vækstmønstre kvantificeret i en Crouzon-musemodel.

Preface

This thesis was prepared at the Section for Image Analysis and Computer Graphics at DTU Compute and submitted to the Technical University of Denmark (DTU) in partial fulfillment of the requirements for acquiring the Ph.D. degree in Applied Mathematics. Selected parts of research work carried out during the Ph.D. period constitutes this thesis totaling four research papers and an introductory part containing an overview of the thesis and background information.

The work was carried out in collaboration with the 3D Craniofacial Image Research Laboratory - an interdisciplinary research unit at School of Dentistry, Faculty of Health Sciences, University of Copenhagen (KU), Denmark. Part of the research was carried out at the Computational Radiology Laboratory, Children's Hospital, Harvard Medical School, Boston, MA, USA under supervision of Professor Simon K. Warfield.

The project was supervised by Professor Rasmus Larsen (DTU), Associate Professor Rasmus R. Paulsen (DTU) and Research engineer Tron A. Darvann (KU) together with the clinical supervisors Associate Professor Nuno V. Hermann (KU) and Professor Sven Kreiborg (KU).

Kgs. Lyngby, January 2013

Signe Strann Thorup

Acknowledgments

I would like to acknowledge several people who put their time, effort and energy into this project. First, I would like to thank all current and former colleagues of the Image Analysis and Computer Graphics Group at DTU Informatics, for providing an inspiring working environment and stimulating various collaborations. I am grateful for all the new friendships I have formed. A special thanks goes to my gang of girls, Line Harder Clemmensen and Stine Harder, for keeping me going and supporting me through, so far, the most turbulent year of my life.

I would also like thank my supervisors, Rasmus Larsen and Rasmus Paulsen, from DTU Informatics and Tron Darvann, Nuno Hermann, and Sven Kreiborg, from 3D-lab, School of Dentistry, University of Copenhagen for their encouraging guidance and supervision throughout the project. Also thanks to Per Larsen at 3D-lab for fruitful discussions over the past years.

Thanks to my friends and family for their help and support, especially during the thesis writing period. I would have been starving and live a less cheerful life without you. Finally, I would like to thank Line Harder Clemmensen, Stine Harder, Mark Lyksborg, Rasmus Ramsbøl Jensen, Jannik Boll Nielsen and Jacob Schack Vestergaard for spending their time reading my thesis in the busy month of December.

Papers included in the thesis

The papers included in the thesis are ordered with respect to chapter number.

Chapter 9 Rasmus R. Jensen, Signe S. Thorup, Rasmus R. Paulsen, Tron A. Darvann, Nuno V. Hermann, Per Larsen, Sven Kreiborg, Rasmus Larsen. *Genus Zero Graph Segmentation: Estimation of Intracranial Volume.* Accepted for SCIA, 2013.
Awarded: Best Student Paper at SCIA 2013.

Chapter 10 Signe S. Thorup, Tron A. Darvann, Nuno V. Hermann, Per Larsen, Rasmus R. Paulsen, , Alex A. Kane, Lun-Jou Lo, Rasmus Larsen, Sven Kreiborg. *A Method for Automated 3D Quantification in CT Scans of Infants with Cleft Lip and Palate-Craniofacial.* Technical report - to be submitted to Cleft Palate Journal.

Chapter 11 Signe S. Thorup, Tron A. Darvann, Nuno V. Hermann, Per Larsen, Hildur Ólafsdóttir, Rasmus R. Paulsen, Alex A. Kane, Dan Govier, Lun-Jou Lo, Sven Kreiborg, Rasmus Larsen. *Dealing with Difficult Deformations: Construction of a Knowledge-based Deformation Atlas.* In Proceedings for SPIE Medical Imaging, 2010.

Chapter 12 Signe S. Thorup, Tron A. Darvann, Hildur Ólafsdóttir, Nuno V. Hermann, Rasmus R. Paulsen, Per Larsen, Chad A. Perlyn, Rasmus Larsen, Sven Kreiborg. *Quantification of Craniofacial Growth in Mice with Craniofacial Dysmorphology Caused by the Crouzon Mutation* $Fgfr2^{C342Y}$. Submitted to Journal of Anatomy 2012.

Abstracts and Papers not Included in the Thesis

[194] Signe S. Thorup, Tron A. Darvann, Nuno V. Hermann, Hildur Ólafsdóttir, Per Larsen, Rasmus R. Paulsen, Rasmus Larsen, Chad A. Perlyn, Sven Kreiborg. *Automatic assessment of craniofacial growth in a mouse model of Crouzon syndrome.* Accepted for the American Cleft Palate - Craniofacial Association's 66th Annual Meeting and Pre-Conference Symposium, 2009.

[195] Signe S. Thorup, Hildur Ólafsdóttir, Tron A. Darvann, Nuno V. Hermann, Per Larsen, Rasmus R. Paulsen, Chad A. Perlyn, Gillian M. Morriss-Kay, Sven Kreiborg, Rasmus Larsen. *Multivariate Analysis of Variance: Finding significant growth in mice with craniofacial dysmorphology caused by the Crouzon mutation.* Accepted for Proceedings of the 8th French-Danish Workshop on Spatial Statistics and Image Analysis in Biology. 2010.

[49] Tron A. Darvann, Nuno V. Hermann, Sune Demant, Per Larsen, Hildur Ólafsdóttir, Signe S. Thorup, Marek Zak, Angelo B. Lipira, Alex A. Kane, Daniel Govier, Helena Schatz, Daniel Rueckert, Sven Kreiborg. *Automated Quantification and Analysis of Facial Asymmetry in Children with Arthritis in the Temporomandibular Joint.* Accepted for International Symposium on Biomedical Imaging, 2011.

[78] Michael S. Hansen, Signe S. Thorup, Simon K. Warfield. *Polyaffine Parametrization of Image Registration based on Geodesic Flows.* Accepted for Mathematical Methods in Biomedical Image Analysis 2012.

Other Material

- Signe S. Thorup, Michael S. Hansen, Tron A. Darvann, Nuno V. Hermann, Per Larsen, Rasmus R. Paulsen, Sven Kreiborg, Rasmus Larsen, Simon. K Warfield. Poster on *2D Single Image Super-Resolution*, Summer school 2011, Bornholm, Denmark.
- Signe S. Thorup, Tron A. Darvann, Nuno V. Hermann, Per Larsen, Hildur Ólafsdóttir, Rasmus R. Paulsen, Alex A. Kane, Dan Govier, Lun-Jou Lo, Sven Kreiborg, Rasmus Larsen. *Poster on Cleft Variation (PCA/SPCA) and Cleft Prediction (linear regression)*, Summer school 2010, Holar, Iceland.

Contents

Part I

Introduction

Chapter 1

Introduction

The skull of a newborn infant is composed of several bones, separated by growth zones, also commonly called sutures. For the skull and brain to grow normally, the sutures must remain open to allow displacement of the bones, forming the adult skull. Several factors can disturb the normal growth process. Conditions associated with disturbed craniofacial development are often referred to as craniofacial anomalies.

3D volumetric images play an integrated and crucial role in diagnosing children with craniofacial anomalies, as detailed views of the internal anatomy are presented. The most common volumetric imaging techniques are Computed Tomography (CT), Cone-Beam Computed Tomography (CBCT) and Magnetic Resonance Imaging (MRI). Currently, the analysis of these images is most often subject to a qualitative inspection by a specialist, while quantitative analysis relies on time-consuming manual or semi-automatic measurements on a limited number of reference points.

The aim of this thesis is to automate the analysis using image processing techniques and thereby increase the level of detail and reproducibility. Moreover, automatic methods would make it possible to increase the amount of data processed. Benefits from a clinical perspective are to understand the abnormal craniofacial shape and growth better by examining the craniofacial anatomy and pathology in detail.

In particular, the thesis focuses on the analysis of craniofacial morphology and growth, including volumetric measurements, and treatment outcomes in three craniofacial anomalies: Crouzon syndrome (in mice), unicoronal synostosis and cleft lip and palate.

1.1 Objectives

The overall goal of the thesis is to investigate and apply image analysis methods to craniofacial images in order to partly or fully answer the following clinical questions:

1. Which changes took place since the last examination? This could be in relation to growth and/or treatment.

2. Which features are typical for a particular type of craniofacial anomaly, and how do they vary in a population?

3. How severely is the craniofacial malformation affecting the individual compared to normal individuals?

In order to answer these questions from a technical point of view it is essential to be able to compare images either from two or more different time points or between patients. This involves creation of correspondence between similar anatomical features which can be obtained by for instance non-rigid image registration. This approach enables one image to be deformed into the coordinate system of another image and, thus, making voxel-wise quantitative comparison possible. From a technical point of view the objectives are the creation and applications of a computational atlas including the analysis of shape from the deformation fields.

1.2 Thesis Overview

This thesis is composed of two parts: First, an introductory part including discussion and conclusion (Chapters 1-7), and second, a contribution part consisting of four papers (Chapters 8-11). The introductory part provides a background and overview of the contributions and basic methods applied in the thesis. The second part includes a selection of papers written during the project period.

Each paper comprises a chapter and a brief description of each of them is given below. Main topics are volume estimation, assessment of variation and modeling change due to growth and surgery. The chapters cover volume estimation of the interior of the skull in Chapter 9, estimation of variation in cleft lip and palate in Chapter 10, modeling changes due to growth and surgery in Chapter 11 and finally, modeling changes due to growth in Chapter 12. The recommended reading guidelines for the thesis are to start with Chapters 2-6. Part II should be read before proceeding to the discussion and conclusion in Chapter 7.

Chapter 9 proposes a method for estimating the intracranial volume in infants with unicoronal synostosis. A Gaussian Mixture Model and graph cuts are used in a fully automatic 3D graph-based method for segmentation of the intracranial volume in CT scans. We reformulate the intracranial volume segmentation problem as an optimal genus 0 segmentation problem in a volumetric graph. The automatic method is validated against a semi-automatic, threshold-based, slice-wise method. Challenges have been related to the holes in the skull e.g. fontanelles, fissures, synchondroses, foramen magnum, and other foramina.

Chapter 10 describes the variation in pre-op CT scans of children with unilateral cleft lip and palate. Non-rigid image registration based on B-splines has been used to model the variation by the means of an atlas as common reference frame. The deformation fields are validated using landmarks. A challenge in this study have been inconsistencies in open/closed mouth, which affects both bone and especially soft tissues.

Chapter 11 proposes a method for bridging pre- and post-op CT images of children with unilateral cleft lip and palate. Non-rigid image registration based on B-splines, has in combination with thin-plate splines, been used to drive a fusion of the soft tissue cleft. The deformation is validated using closest point deformation on extracted surfaces. Challenges in this study have been changes due to growth and the huge change in morphology caused by the lip adhesion procedure.

Chapter 12 proposes a method for modeling growth using non-rigid image registration based on B-splines. Data consist of normal mice and mutant mice with Crouzon syndrome at 4 and 6 weeks of age. The mice were euthanized prior to micro-CT scanning. The non-rigid registration models the growth vectors after removal of translations and rotation. Multivariate analysis of variance, local false discovery rate, surface animations, visualization of growth vectors and linear measurements using landmarks have been used to visualize and quantify growth. Challenges have been related to the cross-sectional material as well as image registration challenges arising from different morphologies i.e. the mouse-types and the age-differences.

Chapter 2

Craniofacial Anatomy and Anomalies

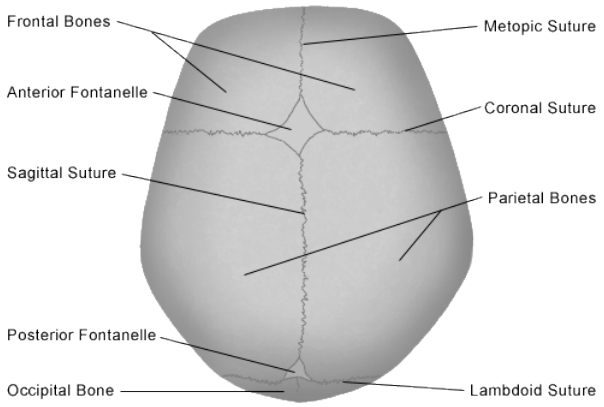
The aim of this chapter is twofold: firstly, to give a brief introduction to craniofacial anatomy in humans and mice, and, secondly, to introduce the craniofacial anomalies studied in this thesis.

2.1 Craniofacial Anatomy in Humans and in Mice

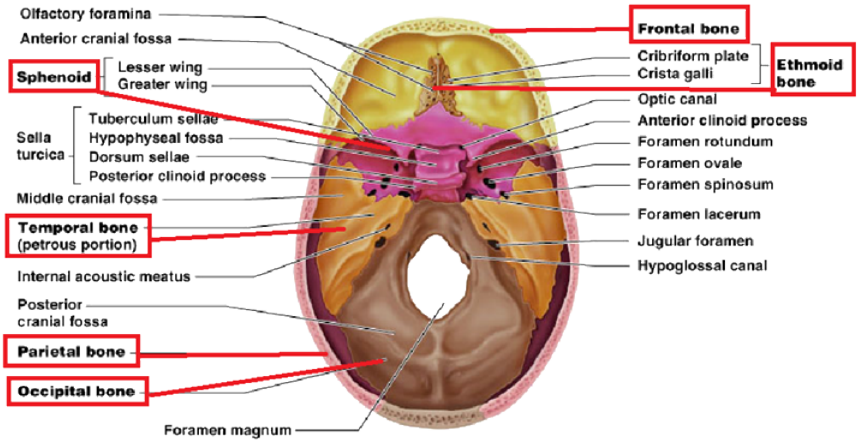
2.1.1 Craniofacial Anatomy in Humans

When we are born the human skull consists of 5 major bones, two *frontal bones*, two *parietal bones* and the *occipital bone*, see Figure 2.1, [185]. The bones are separated by *growth zones* (also known as sutures), which consist of connective tissue. Because of the sutures' flexible substance they allow the skull and brain to grow normally, and the sutures close gradually until adulthood [150]. The major sutures are the *metopic suture*, which extends from the top of the head to the nasal bones and separates the frontal bones; the *sagittal suture*, which extends from the front of the head to the back of the head and separates the parietal bones; the *coronal sutures*, which extend from one ear to the other and separates the frontal bones from the parietal bones; and finally the *lambdoid*

suture, which extends across the back of the head, and separates the parietal bones from the occipital bone. The sutures meet at two soft spots called the *fontanelles*. The *anterior fontanelle* located where the parietal and the frontal bones meet, and the *posterior fontanelle* is located where the parietal bones and the occipital bone meet. Both fontanelles are made of strong membranes, and they close within the child's first two years; the posterior fontanelle is the first to close, typically when the infant is two months old.



(a) Sutures and major bones.



(b) Foramina in the skull base.

Figure 2.1: (a) Schematic figure of a human skull as seen from above with sutures and major bones labeled. Adopted from [131, 185]. (b) shows the various foramina in the skull base, adopted from [130, 185].

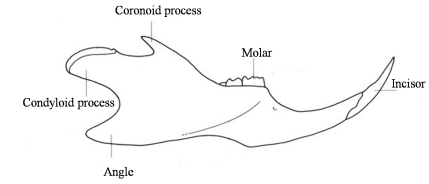
The main biologic functions of sutures are, 1) to act as sites of growth, 2) to be the articulations holding the skull components together, but still allow for the compression and minor movement during childbirth and 3) to absorb mechanical stress and thus protect the tissue where growth takes place. Cranial sutures may prevent separation of calvarial bones due to external forces and trauma because of their interlocking junction. For protective and structural reasons sutures gradually loose the mobility characteristic of the neonate and young infant and become more rigid [150].

At birth, sutures are straight lines of bone edges. Gradually, saw-tooth shaped bony edges develop and become more pronounced. The actual mechanisms resulting in synostosis is not fully understood, but using an animal model Burke et al. [33] proposed a hypothesis for craniofacial suture fusion termed closure by bone bridging. Burke et al. [33] found that *osteoblasts*, cells for formation of new bone tissue, did not extend across the sutural space, and that no *osteoclasts*, cells which remove bone tissue, were found in the sutural space. Hence, with no osteoclasts present there was no mechanism for removal of the microscopic spicules that generate the saw-tooth look of the sutures. Thus, normal mechanical forces cause the spicules to fracture. As a consequence, a suture fuses if a spicule does not fracture, but instead functions as a scaffold, upon which more bone is deposited.
Albright and Byrd [6] and Cohen [39] supported the hypothesis that craniosynostosis begins at one point of the suture, posterior, anterior or in the middle, and gradually spreads along the suture. Thus, only a tiny connection between the spicules will make a whole suture inactive and fusion will follow.

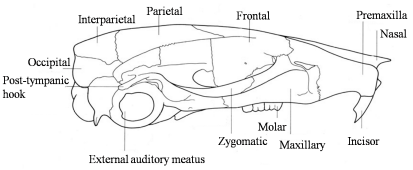
2.1.2 Craniofacial Anatomy in Mice

The mouse skull consists of 6 major bones (Figure 2.2), two *frontal bones*, two *parietal bones*, the *interparietal bone* and the *occipital bone*, see Figure 2.2(d). The major sutures are the *sagittal suture* that runs from the interparietal bone and down to the frontal bones, the *interfrontal suture* that runs from the frontal bones down to the nasal bones, the *coronal sutures* that run from one ear to the other separating the parietal bone from the frontal bones and, finally, the *lambdoid sutures*, which separates the parietal bones from the interparietal and occipital bones i.e. the back of the head, [140, 141]. Figure 2.2(e) illustrates the mouse sutures.

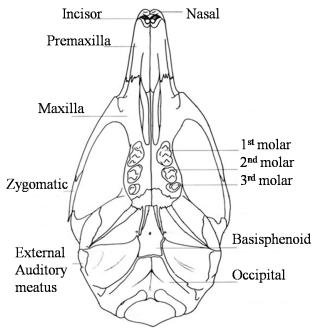
Unlike in humans, where the first suture starts fusing in the first year, all calvarial sutures remain patent in mice, except the posterior part of the interfrontal suture. The interfrontal suture fuses between 25 and 45 days postnatally and fuses from



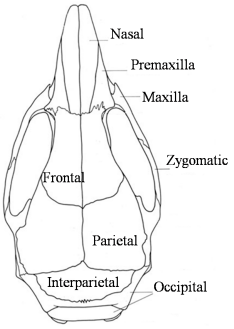
(a) Right lateral view without the mandible



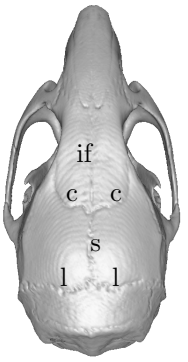
(b) Right lateral view of the mandible



(c) Basal view



(d) Top view



(e) Sutures

Figure 2.2: Schematic drawing of an adult mouse skull (a)-(d) from different angles (Adopted from Cook [41]). (e) Sutures viewed on an adult wild-type skull (6-week). *c* denotes the coronal sutures, *l* the lambdoid sutures, *s* the sagittal suture and *if* the interfrontal suture. Notice that the interfrontal suture is almost closed in this animal. Adopted from Richtsmeier et al. [152, 190].

the anterior part to the posterior [149].

2.2 Craniofacial Anomalies

Sometimes the skull growth is disturbed by several factors leading to abnormal growth and dysmorphology, often with asymmetry of the head and face. Listed below are the most common categories associated with craniofacial malformations:

- *Cleft lip and/or palate.* A congenital malformation characterized by a discontinuity or cleft of the lip and/or the palate. Cleft lip and/or palate can occur as an isolated malformation (non-syndromic), or in association with other congenital malformations (syndromic).
- *Craniosynostosis.* A congenital malformation where the sutures close prematurely, leading to abnormal skull growth. Craniosynostosis can occur as a non-syndromic malformation or it can be part of a genetic syndrome. Syndromic cases include fusion of one or more sutures in the skull with associated physical deviations, e.g. Apert or Crouzon syndrome. The non-syndromic cases include fusion of one or more sutures only, e.g. unicoronal synostosis.
- *Deformational (or positional) plagiocephaly.* Development of an asymmetrical skull shape from repeated pressure to the same area of the head, e.g. by positioning the infant on the same side of the head during sleep.
- *Hemifacial microsomia.* A congenital malformation where one side of the face is underdeveloped, affecting primarily the ear, mouth and jaw. Sometimes, both sides of the face can be affected and the anomaly may involve the skull, as well as the face.

Three different craniofacial anomalies are studied in this thesis: *Crouzon syndrome* in a mouse model, Chapter 12; *unicoronal synostosis*, Chapter 9 and *Cleft lip and palate*, Chapter 10, 11.

2.2.1 Craniosynostosis

This section provides information about the mechanics behind craniosynostosis as well as more details concerning Crouzon syndrome and UCS.

2.2.1.1 Sutures and Bone Growth

As mentioned above, sutures are craniofacial articulations in which two bone ends approximate each other and are united by a thin layer of fibrous tissue [39].

In normal growth, two principles are involved in the postnatal craniofacial growth process: displacement and remodeling of bone. Remodeling serves to adjust for necessary changes in the curvature of individual bones, whereas displacement depends on the presence of normally functioning sutures, which act as areas of growth and adjustment. In cases of premature craniosynostosis the premature fusion of the cranial sutures prevent normal displacement between the bones affected by the fusion, but it is still possible for the affected region to grow due to remodeling or through compensatory growth in patent sutures. This leads to the abnormal growth pattern. Since the timing of suture closure may be random, as well as which suture(s) close first, it is not possible to provide a simple model of the complex changes in craniofacial growth [103].

2.2.1.2 Crouzon Syndrome

(a) Crouzon Syndrome in a human.

(b) Crouzon Syndrome in a mouse model.

(c) Child with UCS.

Figure 2.3: 2.3(a) Child with Crouzon Syndrome (by courtesy of Sven Kreiborg). 2.3(b) Normal mouse to the right, and a Crouzon mouse to the left. 2.3(c) Child with Unicoronal Synostosis (UCS). Adopted from [140, 93]

Crouzon syndrome is characterized by a constellation of premature fusion of the cranial sutures, orbital deformity, maxillary hypoplasia, beaked nose, crowding of teeth, and high arched or cleft palate [140], see Figure 2.3. Usually, the coronal sutures are affected in combination with the sagittal and/or lambdoid sutures. As a consequence of craniosynostosis the head develops an abnormal shape (dysmorphology) due to dysplastic growth but also compensatory growth to accommodate the growing brain [102]. The skull shape depends on the order and the rate of progression of the suture fusions as well as on the number of

sutures involved [150]. In general, the earlier the craniosynostosis takes place and the more involved sutures the greater the effect on skull shape [39, 150].

Treatment of Crouzon syndrome entails both neurosurgery and plastic surgery and usually several procedures have to be carried out. Orbits, midface, lower jaw and deformities in the oral cavity are remodeled and the fused sutures are reopened. Crouzon syndrome is one of the most common syndromes with craniosynostosis with a prevalence of about 1 in 60,000 live births [91].

2.2.1.3 Unicoronal Synostosis

Unicoronal synostosis (UCS) is a congenital craniofacial malformation characterized by the premature fusion of one of the coronal sutures, potentially leading to craniofacial growth disturbances and craniofacial asymmetry with risk of increased intracranial pressure and developmental delays, see Figure 2.3(c). Computed Tomography (CT) scanning is usually performed to confirm the diagnosis and to facilitate surgical treatment planning [150].

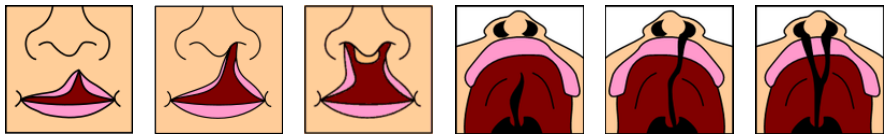
Treatment entails a frontal craniotomy with total forehead reconstruction including the orbital rim area. UCS affects approximately 1 in 10,000 live births [150].

2.2.2 Cleft Lip and Palate

Cleft lip and/or palate is the most common congenital craniofacial malformation, even though inter-racial differences occur [129]. Cleft lip and/or palate can appear as an isolated birth defect or be one of several congenial malformations. If cleft lip and/or palate appears with other malformations it is referred to as *syndromic*, while the isolated cleft is referred to as *non-syndromic* [14, 165]. Within non-syndromic clefts four major categories exist [45]:

1. Cleft lip (CL)
2. Cleft palate (CP)
3. Combined cleft lip and palate (CLP)
4. Atypical clefts

This thesis will focus on complete CLP i.e. the lip, alveolus and palate is clefted and there is no soft tissue bridging [80], see Figure 2.5.



(a) Incomplete cleft lip. (b) Unilateral complete cleft lip. (c) Bilateral complete cleft lip. (d) Incomplete cleft palate. (e) Unilateral complete cleft lip and palate. (f) Bilateral complete cleft lip and palate.

Figure 2.4: Schematic illustrations of cleft classifications. Adopted from [113].

CLP can be uni- (UCLP) or bilateral (BCLP) and it may be complete or incomplete [116], see Figures 2.4. In 90% of CLP-cases, the cleft malformation is unilateral and among these the cleft occurs twice as often on the left side as on the right side [89, 116]. CLP affect 1 in every 700 live births worldwide with wide variability related to ethnicity [129].

2.2.2.1 Definition of Cleft Lip and Palate

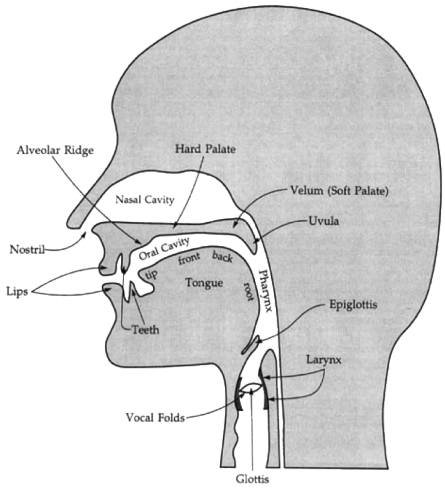


Figure 2.5: Schematic figure of a human mouth in the sagittal plane. Adopted from [188, 185].

Forming a child's head is a complex process. During the first 6-8 weeks of pregnancy where the head of the fetus forms and develops, five tissue lobes emerge; one from the top of the head, two from the cheeks, and two just below

the cheeks. The lobes from the head and cheeks form the upper lip and later in the process the palate, whereas the lower lobes form the lower lip and chin. If the lobes fail to meet and fuse, a cleft occurs [116].

2.2.2.2 Treatment

(a) 6 month old girl before going into surgery for a lip repair.

(b) The same girl, 1 month after the surgery.

Figure 2.6: Girl with UCLP before and after lip adhesion. Adopted from [113].

CLP may seem as a relatively simple anomaly, however, as it interferes with two of the most important means of communication i.e. facial expression/appearance and speech, habilitation presents a major clinical challenge [45]. An example of a girl before and after surgery is provided in Figure 2.6.

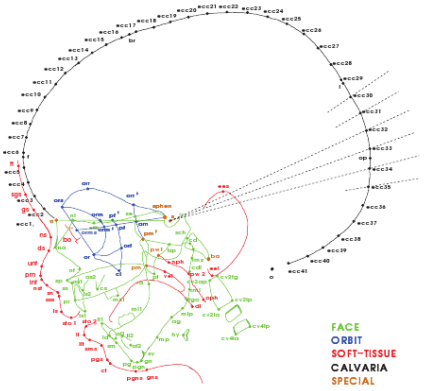
CLP is typically treated by surgical closure of the lip and palate during the first two years of life. The first surgical procedure takes place at about 3-5 months of age and aims at closing the cleft of the lip. Many surgical approaches exist, but the so-called lip adhesion is a common way of closing the cleft by stretching the soft tissue of the lip across the cleft. In the case of an additional clefting of the bone (teeth ridge and/or hard palate), the forces from the lip tissue stretching will also lead to a slight narrowing of the bony cleft during the time from lip adhesion to the time of surgical closure of the palate which takes place approximately a year later [116]. As surgical repair is usually carried out early in life there is a great interest in assessing the surgical outcome [204, 17].

Chapter 3

Image Acquisition in Craniofacial Research

This chapter aims at providing an introduction to the most used image modalities within the orthodontic practice and craniofacial research. Imaging techniques

(a)

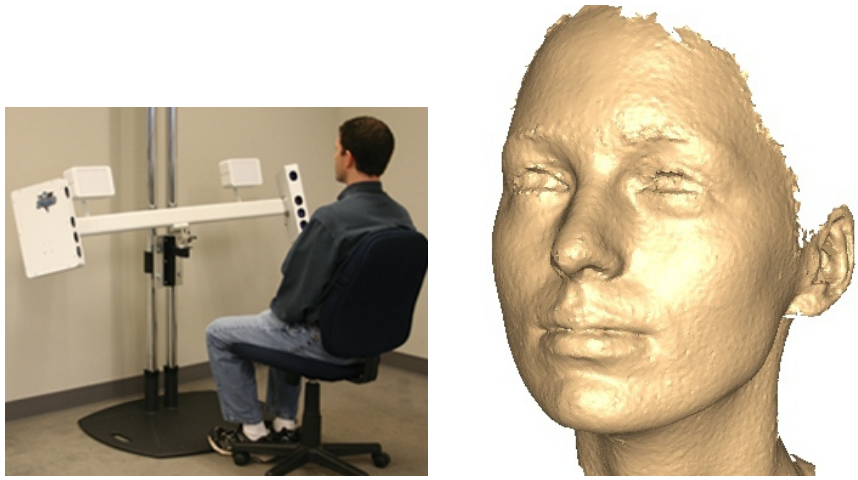


(b)

Figure 3.1: A modern cephalostat is shown in (a), [51]. (b) depict a cephalogram from a lateral X-ray with landmarks and their names, adopted from [80].

for craniofacial research have been available for over a century with Roentgen's discovery of X-rays in 1895 [13]. The next epoch event within craniofacial research occurred in 1931 when roentgencephalometry was introduced simultaneously by Broadbent [31] in the USA and Hofrath [83] in Germany. Broadbent standardized the cephalometric technique by using an X-ray machine and a head holder called a cephalostat or cephalometer [13], see Figure (a). Today roentgencephalometry is still the most widely used method in orthodontic practice and research, e.g. to assess craniofacial morphology, and growth, and to evaluate the outcome of treatment regarding the dentition and the facial skeleton [110, 68, 167, 179]. Analysis of the cephalometric radiograph (also known as cephalogram, see Figure (a)) includes identifying landmarks, and measuring distances and angles between landmark locations [102]. Cephalograms are most often acquired in the lateral and frontal views, but more advanced versions including the axial view for three-dimensional analysis have been developed (e.g. [104, 182]).

In the 1970's and 1980's growing awareness of the aesthetic outcome for the patients lead to a (re)introduction of soft tissue evaluation through (2D) photographs, e.g. [12, 199]. Today 3D imaging techniques provide a helpful tool in orthodontics, e.g. 3D photography, MRI, CT and CBCT are used more and more [143].

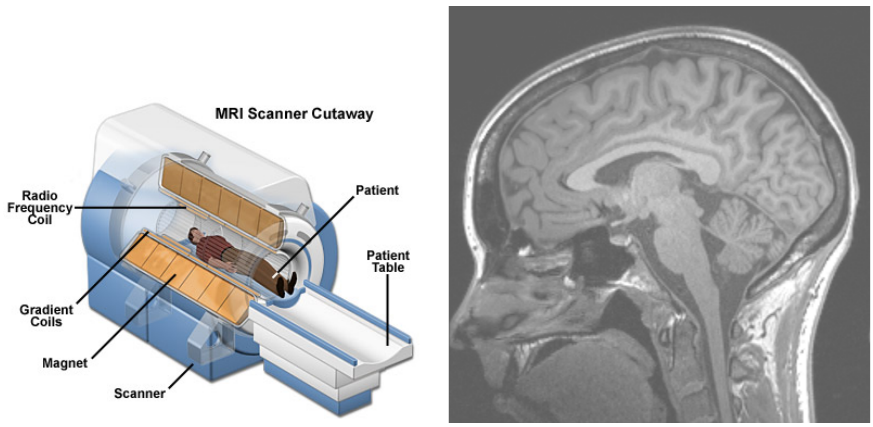


(a) (b)

Figure 3.2: 3D surface scanner setup with two camera pods is shown in (a), [1]. (b) shows an example surface scan from a 3D surface scanner.

3D photography also known as (stereo)photogrammetry capture photo-realistic 3D textured surfaces of e.g. objects or faces with metric accuracy. The setup

consists of two or more camera-pods in a stereo setting that acquires images in milliseconds, e.g. as the setup in Figure (a). From the 2D images a 3D surface is reconstructed with texture information [143]. An example of a 3D surface can be found in Figure (b).

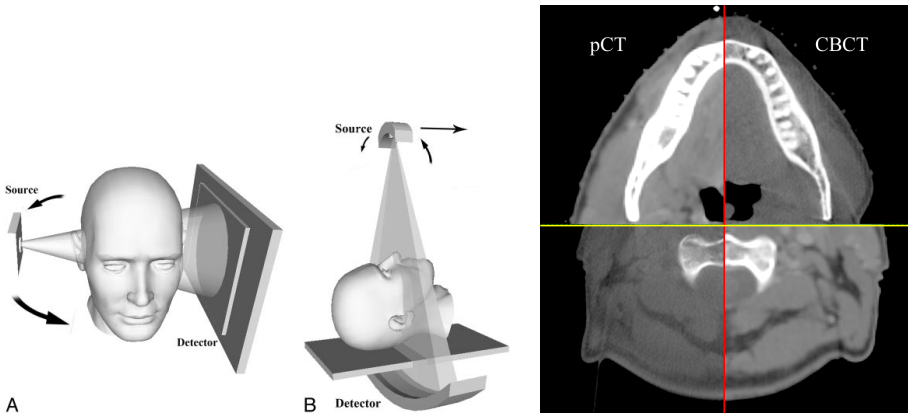


(a) (b)

Figure 3.3: (a) depicts a MRI scanner, [4]. (b) shows a sagittal T1-weighted MR slice.

MRI uses a powerful magnetic field along one direction to align the spin of the hydrogen nuclei in the body. A magnetic pulse is then induced in order to push the spins away from equilibrium. The spins will return to equilibrium due to near-collisions between nuclei; this process is called relaxation. The nuclei emit radio waves during relaxation and this signal is measured by a receiver coil. The contrast in MRI occurs because of differences in relaxation time between tissues. A large number of image sequences exist and they are used in order to achieve a desired signal difference between tissues. The spatial position of a measured signal is determined by inserting a field gradient along the three axes of the patient []. A field gradient along one axis results in a change of the resonance frequency along the axis. If a pulse with a certain frequency is induced, only the nuclei with that resonance frequency will be turned away from equilibrium, resulting in slice detection. The more complex K-space representation is used in order to obtain the spatial position within a slice. Thus, multiple 2D images are produced, that can be reconstructed into 3D volumes, a single slice is sown in Figure (b). The main advantage of MRI is the lack of ionizing radiation and information about the inner soft tissue layers. The patient is usually positioned in the supine position in the scanner [79], Figure (a).

In CT, an X-ray tube is rotated around a given object, while multiple 2D detectors on the opposite side of the source measure the attenuations of the fan-shaped rays. As different tissues have different densities and attenuations, an image of the interior of an object can be reconstructed by estimating the spatial distribution of the linear attenuation coefficient. In practice, this means to recover the linear attenuation coefficients at a finite number of spatial positions in a regular grid [146, 184]. The patient is usually positioned in the supine position in the scanner. Cone-beam CT (CBCT) works similar to CT, but the beam is cone-shaped while CT often has a fan shaped beam. The patient is, in most CBCT scanners, positioned upright in a natural position, and a single large 2D detector is used. The radiation dose is much lower for the CBCT than for multi-detector CT [184, 146, 143]. In Figure (b) an image comparing CT and CBCT is shown.



(a) (b)

Figure 3.4: (a) depicts a cone-shaped CBCT beam (left), and a fan-shaped CT beam (right). The arrows indicate the movement of the X-ray source, [174]. (b) CT (planning CT) and CBCT (Courtesy of Rikke Eiland)

Although roentgencephalometry has advantages compared to CT and CBCT such as the low cost and low radiation dose, the 3D imaging techniques have a huge advantage compared to the conventional radiographs not only because of the 3D details but also due to the elimination of magnification errors, image distortion and errors due to superimposition of anatomical structures [143].
3D photogrammetry's advantages are the low cost w.r.t. acquisitions, accurate surface representation, lack of ionizing radiation and moving artifacts, while the disadvantages are the potential need for daily calibration, poor precision on shiny surfaces e.g. teeth or glasses, difficulties with hair and undercut areas e.g. ears, subnasal area, and it is only suitable for soft tissue [143]. MRI, CT, and

CBCT are all able to visualize soft tissue and bone while providing the patient in 3D for a more accurate examination. MRI might be preferable due to the absence of ionizing radiation, but it is expensive, has long acquisition time with the risk of movement, the position of the patient w.r.t. soft tissue shape, and finally, the quality of the skeletal data is not acceptable for planning surgery etc. Both CT and CBCT are able to display the facial skeleton accurately [143].

In a number of clinical areas, CBCT supersedes conventional CT due to lower costs, the upright position for better and more natural visualization of soft tissue, as well as the lower radiation dose. Especially in dentistry, CBCT is preferred to CT since CT generates streak artifacts from orthodontic brackets and metal restorations; and acrylic resin fillings are similar in gray scale to soft tissue [143]; disadvantages are, however, more scattering and reduced contrast in the CBCT images complementing to the lower radiation dose [143, 64]. Even though CBCT has a higher dose compared to a single radiograph; several radiographs will exceed CBCT w.r.t. dose including the fact that CBCT will provide a more detailed and true study of the patient [209, 143].

Chapter 4

Data

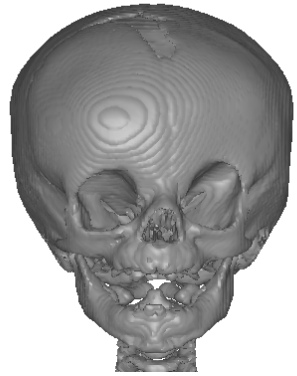
This Chapter introduces the data used in this thesis: UCS, UCLP and Crouzon syndrome in mice.

4.1 Children with Unicoronal Synostosis

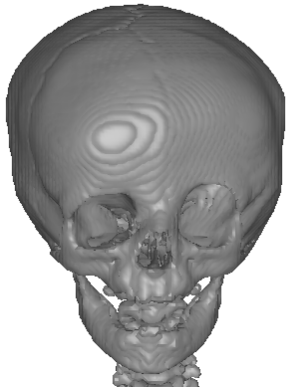
The data consist of pre-op CT head scans of 15 children diagnosed with UCS (either left- or right-sided, termed LUCS and RUCS, respectively). All scans were acquired at Copenhagen University Hospital, Rigshospitalet, except for one which was acquired at Helsinki University Central Hospital. All scans were obtained at 512 x 512 pixels in-plane size. However the number of slices vary between 167 and 350. Table 8.2 in Appendix 8.2 provides information about diagnosis, age at scan time, voxel resolution, slice number in the z-direction (axial) and intensity range. Examples of the data are illustrated in Figure 4.1.

Before the segmentation was applied, two pre-processing steps were applied: firstly, noise (e.g. scan bed, scan tube and headrest) was removed using thresholding and dilation, and, secondly, the volumes were resampled and interpolated in the axial direction to create isotropic voxels. In Chapter 9, the focus is on estimating the intracranial volume (ICV). Children with premature fusion of one or more sutures in the calvaria may have reduced ICV, leading to risk of

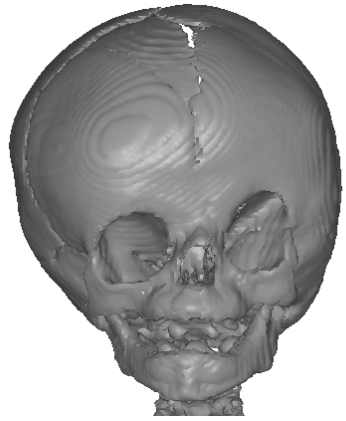
(a) Patient #7: Soft tissue.



(b) Patient #7: Bone tissue.



(c) Patient #9: Bone tissue.

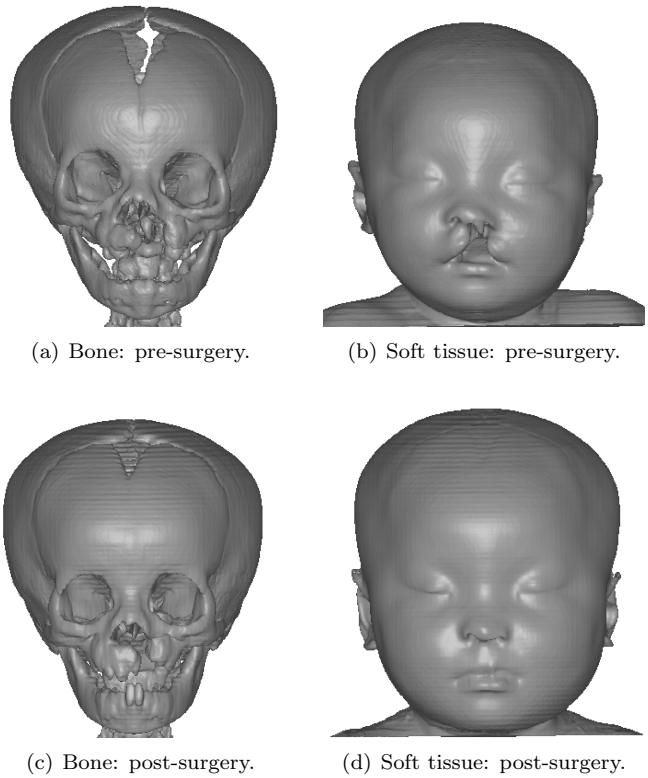


(d) Patient #10: Bone tissue.

Figure 4.1: Iso-surface examples extracted from the UCS data using Marching Cubes [114]. Notice the large unfused fontanelle for patient #10.

increased intracranial pressure [205]. The aim was to develop a method for ICV estimation. Challenges have been related to the holes in the skull e.g. fontanelles, fissures, synchondroses, foramen magnum, and other foramina.

4.2 Children with Unilateral Cleft Lip and Palate



(a) Bone: pre-surgery. (b) Soft tissue: pre-surgery.

(c) Bone: post-surgery. (d) Soft tissue: post-surgery.

Figure 4.2: An iso-surface example extracted from the UCLP data using Marching Cubes [114]. Patient #21 is shown before and after surgery.

CT scans of 23 Taiwanese infants[1] with UCLP were obtained in a clinical context at Chang Gung Memorial Hospital, National Yang Ming University, Taipei, Taiwan. All infants were born between 1997 and 1999, and the use of CT was approved by a Taiwanese Ethical Committee. The infants were scanned before

[1]Five of the images were later discarded due to scanning artifacts.

lip repair at the age of approximately 3 months and again before palate repair (i.e. after lip repair) at approximately 12 months. The data were reformatted in Taiwan w.r.t. intensity range (conversion to 8 bit data), voxel size (0.5 mm) and a number of $340 \times 340 \times 400$ voxels, see Figure 4.2 for an example. See Table 8.3 in Appendix 8.3 for more information about excluded data, intensity range and age range. In UCLP, the cleft can be either to the right side or to the left side – a left side cleft occurs twice as often as a right side cleft [89, 116]. Assuming that it is the same mechanisms that cause the clefts, irrespective of side, all right-sided clefts were mirrored to become left-sided clefts. Pre-processing of the data, thus, included, mirroring of the data, orientation according to the mid-sagittal plane, and identification of all background voxels outside of the subject.

Chapter 10 and 11 both use the UCLP data. Chapter 10 aims at quantifying cleft variation in the pre-surgery patients, while Chapter 11 aims at creating point correspondence between pre- and post-op images. Both Chapters employ the use of an atlas representing the average morphology. Challenges in these studies have been in relation to variation in the dataset e.g. because of plagiocephaly (Chapter 10), while changes due to growth and the huge change in morphology caused by the lip adhesion procedure caused problems in Chapter 11.

4.3 Mouse Model of Crouzon Syndrome

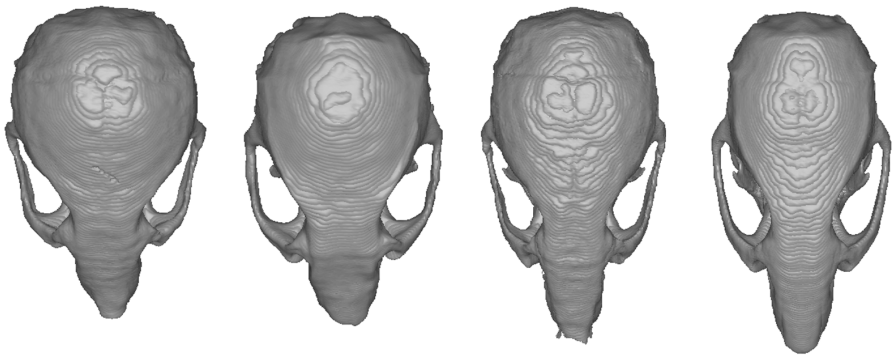


Figure 4.3: Four example mice (from left to right): 4-week-old Crouzon mouse; 6-week-old Crouzon mouse; 4-week-old normal mouse; 6-week-old normal mouse. The iso-surfaces were extracted using Marching Cubes [114].

Micro-CT scans of 4-week-old mice (five specimens) and 6-week-old mice (ten

specimens) with Crouzon syndrome ($Fgfr2^{C342Y/+}$) were compared to similar control groups of wild-type mice. The production of the $Fgfr2^{C342Y/+}$ mutant mouse (Crouzon mouse) was carried out as described by Eswarakumar et al. [62]. All procedures were in agreement with the United Kingdom Animals (Scientific Procedures) Act, guidelines of the Home Office, and regulations of the University of Oxford. Mutant mice of breeding age were determined by phenotype. Female $Fgfr2^{C342Y/+}$ mice were bred with males genetically determined for the same mutation.

For micro-CT scanning, the 15 wild-type (normal) and 15 $Fgfr2^{C342Y/+}$ specimens at 4 and 6 weeks of age (28 and 42 days), respectively, were euthanized by CO_2 asphyxiation, and the skulls were extracted by removing all soft tissue. The mice were sealed in conical tubes and shipped to the micro-CT imaging facility at the University of Utah, Salt Lake City, USA. 3D volumes of the skull of size $480 \times 480 \times 720$ voxels were obtained at approximately $46\mu m \times 46\mu m \times 46\mu m$ resolution per voxel using a General Electric Medical Systems EVS-RS9 Micro-CT scanner.

At the age of 4 weeks, the coronal sutures were fused; there was a partial fusion of the lambdoid sutures; and the sagittal suture was partially separable. By the age of 6 weeks the sagittal suture was completely obliterated in 70% of the mice [140], see Appendix 8. Prior to image analysis, the neck part, hyoid bone and scanning artifacts were removed. Figure 4.3 shows an example of the two mouse types.

In Chapter 12, the focus is on modeling craniofacial growth in a Crouzon mouse model. The aim was to use non-rigid image registration to model the growth in each point on the mouse head. An atlas representing the average craniofacial anatomy of the 6-week-old normal mice was created using non-rigid image registration. The atlas was used to establish a common coordinate system for all mice. Challenges have been related to the nature of the cross-sectional material as there is no correspondence in time between the mice. Moreover, image registration challenges arose from the different morphologies, i.e. the mouse-types and the age-differences.

Chapter 5

Craniofacial Morphology and Growth - From Manual to Automatic Analysis

The analysis of shape has been a central element of much biological research. Morphometrics is the branch within morphology concerning quantitative description, analysis, and interpretation of shape and shape variation. This chapter gives a brief review of two topics relevant to this thesis, namely intracranial volume estimation and landmark-based morphometrics [122].

5.1 Intracranial Volume Estimation

Brain growth drives skull growth during childhood [176, 98]. At about 20 years of age, the volume of the brain starts to decrease, while it is presumed that the ICV remains constant [206]. Knowledge of the cranial cavity volume may be of importance to study as well as comparing the crania across populations w.r.t. e.g. medical, racial, geographic, ethnic, or dietary differences [120]. For instance, in population studies ICV is considered a more accurate measure of mature brain volume than head size [206]. By combining measures of the ICV with measures of

brain volume, investigators can infer how much reduction in volume has occurred since brain volume was at its peak [88, 98]. Hence, the ICV provides a more stable and accurate normalization factor for estimating volumetric changes at the onset of a disease [58]. Through time, several methods have been proposed

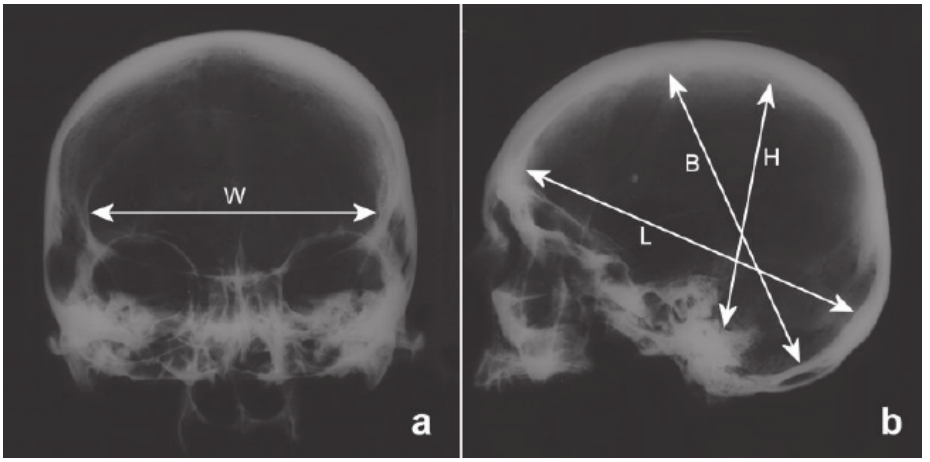


Figure 5.1: Two standard anteroposterior (a) and lateral (b) cephalograms of a skull. Internal length (L), internal height (H), diameter from bregma to posterior cranial fossa (B), and width (W). From [168].

for assessing the ICV, see [120] for more details on accuracy etc. The first attempts to estimate the ICV was carried out on dry skulls either by linear measurements (anthropometry[1]) (measured on dry skulls and in living by Lee and Pearson, 1901, [108]), packing or filling methods (manual or using a machine for the filling) or later on cephalometric radiographs. The linear measurements involved measuring the width, breadth and height of the skull [108], while the packing or filling methods used materials like bullets, sand, paraffin wax, various seeds e.g. mustard, to pour into a skull with all foramina (except the foramen magnum) plugged with wax or cotton, or by the use of water-filled balloons [118, 84, 186, 144, 164]. The best measurements were Uspenskii's [164] and had a maximum difference of 7 cm^3, when compared to a hollow endocast of the same cranium. In comparison, the millet seed method had 65.4 cm^3 difference. A gold standard study by Sahin et al. [168] aimed at comparing the accuracy and the reliability of anthropometry, cephalometry, point-counting, and planimetry[2] methods for the estimation of ICV using dry skulls. However, no automatic methods were used in the study.

[1]Linear measurements such as length, width, and height. Various formulas exists for calculating the ICV from these variables, see e.g. [120]

[2]The boundaries of the intracranial cavity were manually traced slice-wise.

As mentioned above, estimations carried out on living subjects began with linear measurements [108]. With the entry of the X-ray machine, volume estimation could be made on internal measures instead of external, see Figure 5.1. However, the measures were still limited to height, width and length, but now the internal skull width could be incorporated in the formulas. In 1977 a large study on ICV estimation was conducted on 1058 Caucasians in the age group 7 days to 20 years using X-rays and linear measurements [50]. Intracranial volume

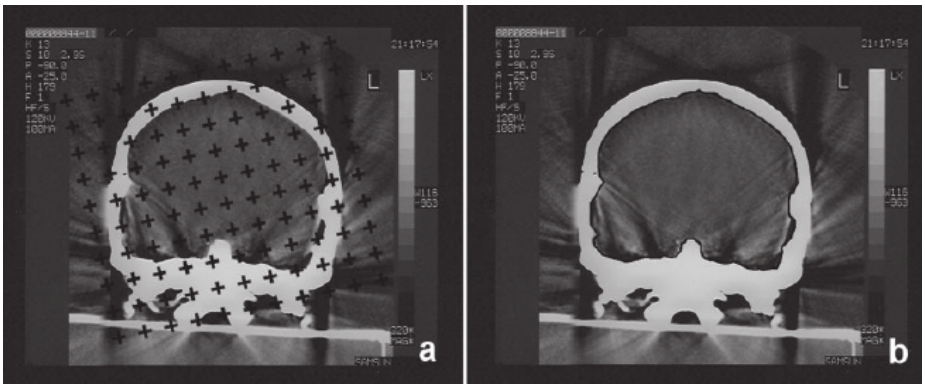


Figure 5.2: A computed tomography scan in a coronal section of a dry skull. In (a) a transparent square grid test system has been superimposed, randomly covering the entire image frame to estimate the section cut surface area. Point counting is a slice-wise, pseudo-integral method where the area of each slice is calculated from a grid randomly superimposed on each slice. After each superimposition, the number of test points hitting the structure of interest on the sections is counted, and the volume of the structure is estimated by multiplying section thickness, total number of points, and the representing area per point in the grid. (b) Inner boundaries of the skull have been manually traced on the image to calculate the section cut surface area using the planimetry method. From [168].

measurements using CT scans have already repeatedly been reported in the literature [171, 69, 117, 2, 66, 71, 203, 145, 175, 3, 7, 8, 121, 208]. Most attempts to determine the ICV were carried out using manual or semi-automatic planimetry [69, 117, 2, 66, 71, 7, 8, 3]. However, manual delineation of the intracranial cavity boundaries is a tedious and labor-intensive process. Other studies have performed volumetric measurements using intensity thresholding and segmentation [171, 203]. The reliability of these measurements is based on the suitable selection of the intensity threshold value to define the intracranial space. Point-counting has the advantage of not depending on the user's skill in delineating the boundaries of the structure of interest [121, 168, 3], but the points need to be counted manually, see Figure 5.2 for a visualization of point counting and planimetry. More sophisticated approaches enabling the automatic segmentation of the

region of interest have also been proposed [145, 175]. Both studies reported that operator intervention was necessary to manually trace the intracranial cavity borders on CT sections, including skull base foramina. Finally, Yasuda et al. [208] proposed an automatic method for ICV estimation using thresholding, morphological operations and contour extraction in crania before and after craniotomy. However, later in the paper they state that the method for post-surgery analysis is 'almost' automatic. This paper will be discussed in Chapter 7. In Chapter 9, we proposed a method for automatic estimation of the ICV using image segmentation involving a Gaussian Mixture Model and graph cuts.

5.2 Landmark-based Analysis

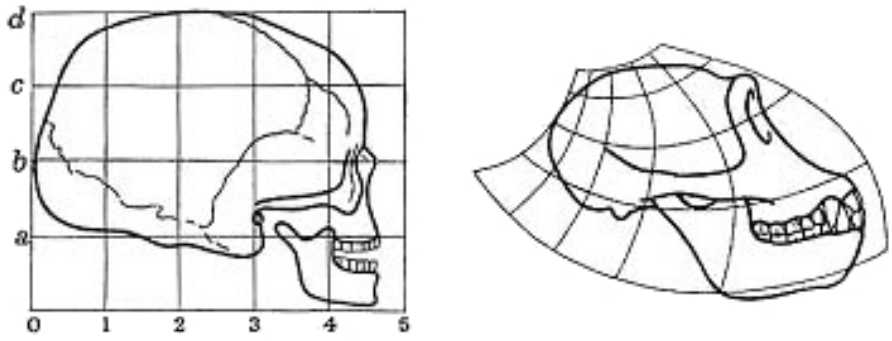


Figure 5.3: Comparison of the shape of a human skull (left) and a chimpanzee skull (right). The hand drawn deformation grids represent the correspondence to a human skull. Adopted from [191].

An early inspiration for morphometrics was D'Arcy Thompson's famous work, *On Growth and Form*, where differences in the (craniofacial) anatomy was examined manually in 2D [191], see Figure 5.3. Traditional morphometrics and the application of multivariate statistical methods to sets of variables was initiated in the 1960's and 1970's [5]. Variables can be linear distance measures such as lengths and widths, but also angles, counts and ratios measured between certain landmarks [157]. Examples of techniques are Principal Components Analysis (PCA), factor analysis, Canonical Correlation Analysis, discriminant functions and generalized distances [157, 5]. Many studies are concerned with allometry (changes in shape wrt. changes in size) and size correction to enable the study of the shape differences. In the combination of quantitative morphology and multivariate statistics several problems remained: First, many methods for size correction were proposed, but there was little agreement on which method

to use, which, as a consequence, lead to slightly different results. Second, the consistency of the linear measurements was difficult to assess, as a consequence of the many distances lacking a golden standard for annotated landmarks. Third, the same set of distance measures can be obtained from two different shapes because the curvature was not included in the data. For instance, the maximum width and height of an ellipse and a teardrop can be the same even though the shape is different. Finally, since the linear measurements are generally insufficient for reconstructing the geometry, some aspects of shape are lost [5].

With the geometric morphometric revolution, biological research started moving from the traditional morphometrics. Outline methods were the first to emerge, and once more no common method was advocated for. Some methods outlined an object from a center point using equal-spaced radii, others used changes in the angle of tangents etc. Furthermore, the shapes had to be relatively simple. As a result the methods worked, but their outcomes were not comparable and no golden standard method could be picked [5]. A consistent methodology was established with the biological identifiable landmarks. Direct analysis of the landmark would prove fatal, but by removing non-shape variation, such as position, orientation and scale, the shape variables could be statistically compared and visualized. Superimposition methods eliminate non-shape variation, and with Generalized Procrustes Alignment (GPA) morphometrics got a common method for non-shape variation removal [5]. Other popular methods without non-shape variation are Euclidean Distance Matrix Analysis (EDMA)[109], Finite Element Scaling Analysis (FESA) [151], methods based on interior angles e.g. [147] and thin-plate splines [56, 23]. However, controversies exists w.r.t. whether inter-landmark distances and angles are biased and have large errors in their mean shape [109, 151, 147, 154, 155, 156]. Also, Walker [202] documented some of the problems in estimating the pattern of variances and covariances of landmarks after a superimposition. However, these problems do not invalidate overall tests for shape differences. Existing methods for estimating these variances and covariances are unsatisfactory because they can show unequal variances and strong covariances within and between landmarks when the actual variation is isotropic. The best advice regarding using landmark-based analysis is that of [153] who suggested performing simulations using the mean consensus configuration as a population mean to make sure that the patterns of variation are not just artifacts of the superimposition procedure.

Today, modern image registration is able to provide 3D deformation grids fully automatically. Still, 3D shape variation using the sparse landmark representations mentioned above is popular as opposed to dense point-correspondences. One example of dense-correspondences is the work of Bugaighis et al. which applied shape modeling to facial 3D photogrammetry data for the classification of different groups of children with clefts [32]. In [9], a method for 3D surface registration of mandibles is applied to growth modeling. Among the pioneers regarding full

volumetric images such as CT and the application of non-rigid registration are Christensen et al. [38] with the creation of a computational cranial atlas. Other examples are, e.g. evaluation of the mandible and the maxilla after surgery [36], and craniofacial mouse atlases [133].
This thesis further adds to the type of applications by including atlas-based variation in UCLP, Chapter 10; atlas-based point correspondences for linking pre-surgery UCLP to post-surgery UCLP, Chapter 11; and finally, quantification of craniofacial growth in mice, Chapter 9.

Chapter 6

Methods

The aim of this chapter is to provide an overview of the methods applied in this thesis. Surface extraction was used to visualize etc. the volumetric data in Chapters 10-12. Image registration is important as it ensures cross subject correspondence, a property used in Chapters 10-12, to examine growth, variation, and changes due to surgery. Multivariate statistical methods are useful when analyzing multiple findings in image analysis in order to test for statistical significance, Chapter 12. In Chapter 9 the following methods were used: Graph-cut is a technique useful for segmenting the skull in a highly efficient manner compared to semi-automatic methods currently used in the clinic, where manual editing often is necessary. Mesh generation constructing a graph for graph cuts. A mixture of Gaussians was employed to act as a prior in image segmentation for graph cuts.

6.1 Surface Extraction

Iso-surface extraction is an essential technique for analysing, visualizing and landmarking medical volumetric images. An iso-surface is a surface that connects all voxels, within a 6-voxel neighborhood, with the same iso-value (or threshold) assigned. Different strategies exist on how to treat each voxel, when creating a mesh from the 3D volume [184]. In this thesis we used Marching Cubes, [114]

for extracting e.g. skin and bone surfaces, see the various data representations in Section 4.

6.2 Image Registration

The goal of image registration is to uncover the optimal transformation $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$, which transforms a source image A such that it becomes similar to a target image B. If we denote the transformed image by $A(T(x))$, the optimal transformation should ensure that $B \approx A(T(x))$. Image registration consists of the following elements

- The geometrical transformation required to transform the source to the target image.
- The similarity measure that describes the quality of the registration.
- The optimization algorithm that controls or determines the parameters of the geometrical transformations to maximize similarity.
- The regularization term securing that only reasonable transformations are obtained.

Each of the items listed is a research area in its own right. A great number of different algorithms and methods for each discipline exists. Throughout the thesis the intensity-based image registration framework, IRTK, developed by Rueckert et al. [163, 170, 162] was employed along with landmark-based registration in the form of volumetric thin-plate splines incorporated in the framework by Darvann [48].

6.2.1 Landmark-based Registration

Formally, a landmark point can be defined as a point of correspondence, that matches objects within and between populations under examination, on each object.We will begin by considering image registration in cases where the similarity measure is defined in relation to a set of corresponding landmarks in a pair of images.

Thin-plate splines (TPS) were formulated by Duchon [56] and Meinguet [124] for surface interpolation of scattered data. The idea of TPS is to model the behavior of infinitely thin metal plates, when forced through 3D points. Metal plates forced through specific points will exhibit minimum bending energy. The

mathematical modeling of metal plates will thus give the interpolation between points with minimum bending energy. The formulation of TPS carries over to N dimensional space. This section deals with the 3D TPS. The TPS-based geometrical transformation is usually applied in situations where the warp field is to be determined from given correspondences between a smaller number (< 1000) of landmark points in two volumes [107].

Assume that two corresponding point sets have been extracted from a pair of images, i.e. $\bar{\boldsymbol{x}}_i, \bar{\boldsymbol{y}}_i \in \mathbb{R}^3$, $i = 1, ..., N$. The $\bar{\boldsymbol{x}}_i$ coordinates belong to the space of the source image A and the $\bar{\boldsymbol{y}}_i$ coordinates belong to the space of the target image B. We aim at determining the parameters $\bar{\boldsymbol{w}} \in \mathbb{R}^p$, where p is the number of points, of a geometrical transformation $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ that map the coordinates of the source image to the coordinates of the target image such that the sum of squared distances is minimized, i.e.

$$f(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \lambda) = \arg\min_{w} \sum_{i=1}^{N} \|T(\bar{\boldsymbol{x}}; \bar{\boldsymbol{w}}) - \bar{\boldsymbol{y}}_i\|^2 + \lambda J(f) \tag{6.1}$$

where the bending energy (regularizer), $J(f)$, is a function for the curvature of f

$$\begin{aligned} J(f) = \iiint & \left(\frac{\partial^2 f}{\partial x_1^2}\right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2}\right)^2 + \left(\frac{\partial^2 f}{\partial x_3^2}\right)^2 \\ & + 2\left(\frac{\partial^2 f}{\partial x_1 x_2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x_1 x_3}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x_2 x_3}\right)^2 dx_1 dx_2 dx_3. \end{aligned} \tag{6.2}$$

The parameter λ governs the degree of smoothing. If $\lambda = 0$ the TPS is interpolating instead of approximating the $\bar{\boldsymbol{y}}_i$'s. At higher λ values the surface becomes more and more smooth since curvature is penalized [75]. When $\lambda \to \infty$, the TPS surface is reduced to a least square fitted plane, since a plane has bending energy $= 0$.

Modeling each coordinate k of the geometrical transformation as a smoothing TPS transformation; it can be shown [201, 72] that the minimization of f is of the form

$$f(x) = \beta_0^k + (\beta_1^k)^T x + \sum_{i=1}^{N} \alpha_i^k \eta_3(||x - x_i||), \tag{6.3}$$

where

$$\eta_3(r) = \|r\|^3 \tag{6.4}$$

r is the pairwise distance between the two point sets $\bar{\boldsymbol{x}}_i$ and $\bar{\boldsymbol{y}}_i$. The following

constraints apply on α_i

$$\sum_{i=1}^{N} \alpha_i = \sum_{i=1}^{N} \alpha_i x_{i.1} = \sum_{i=1}^{N} \alpha_i x_{i.2} = \sum_{i=1}^{N} \alpha_i x_{i.3} = 0. \tag{6.5}$$

The α_i ensures that the $J(f)$ function is finite. To solve Equation 6.3, the formulas above can be written in matrix form:

$$\boldsymbol{P} = \begin{bmatrix} 1 & ... & 1 \\ \bar{\boldsymbol{x}}_{\mathbf{1}} & ... & \bar{\boldsymbol{x}}_{\boldsymbol{N}} \end{bmatrix}, \qquad 4 \times N \tag{6.6}$$

and

$$\boldsymbol{K_{ij}} = \eta_3(||\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j||), \qquad N \times N \tag{6.7}$$

with the matrices $\boldsymbol{P}$ and $\boldsymbol{K_{ij}}$ the system of equations can be written as

$$\begin{bmatrix} \boldsymbol{K} + \lambda \boldsymbol{I} & \boldsymbol{P}^T \\ \boldsymbol{P} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\alpha}} \\ \bar{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0} \end{bmatrix} \tag{6.8}$$

where $\bar{\boldsymbol{y}} = [\bar{\boldsymbol{y}}_1, ..., \bar{\boldsymbol{y}}_N]^T$, $\bar{\boldsymbol{\alpha}} = [\alpha_1, ..., \alpha_N]^T$ and $\bar{\boldsymbol{\beta}} = [\beta_0; \beta_1]$. This linear system of equations is solved with respect to $\bar{\boldsymbol{\alpha}}$ and $\bar{\boldsymbol{\beta}}$. An estimate of the TPS at the location x can now be calculated using Equation 6.3.

Since the TPS transform is only exact for the anatomical landmark locations, the vertices of the template image will not necessarily lie on the surface of the target shape.

6.2.2 Intensity-based Registration

Rueckert et al. developed a framework based on a transformation T, which is a combination of a global transformation and a local transformation. The transformation can be written as:

$$T(x, y, z) = T_{\text{global}}(x, y, z) + T_{\text{local}}(x, y, z), \tag{6.9}$$

where the global motion describes the overall motion, and the local motion at each point is modeled by 3D B-splines. The optimal transformation T can be found by minimizing a cost function. The cost function compromises two goals: 1) The (dis)similarity measure measures the voxel similarity between image pairs. A good registration has a large similarity score. 2) The smoothness term, also known as the *regularizer*, is applied to ensure a valid transformation without folding.

6.2.2.1 Global Transformation Model

The global model describes the overall motion. One of the simplest transformation models is the rigid transform, which in 3D is parametrized by 6 degrees of freedom (rotations and translations). A more general model is the affine transformation, which have 6 extra degrees of freedom (scale and shearing). The affine transformation can be written as

$$T_{\text{global}}(x, y, z) = \boldsymbol{U}\boldsymbol{x} + \boldsymbol{t} = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \tag{6.10}$$

where $\boldsymbol{x}$ is a vector of pixel/voxel coordinates (x, y, z), $\boldsymbol{t}$ is the translation vector (t_x, t_y, t_z) and $\boldsymbol{U}$ is a matrix describing the rest of the parameters i.e. rotation, (anisotropic) scaling, and shearing [183, 75]. The rigid transformation can be written in a similar way including only rotation and translation i.e. $\boldsymbol{U} = \boldsymbol{R_1}\boldsymbol{R_2}\boldsymbol{R_3}$, where $\boldsymbol{R_1}$, $\boldsymbol{R_2}$ and $\boldsymbol{R_3}$ specifies rotation around each of the three coordinate axes.

6.2.2.2 Local Transformation Model

In order to capture the local differences in the images, rigid or affine transformations are not sufficient. The term non-rigid transformation will be used on equal terms as local transformation, and refers to a transformation with more degrees of freedom (DOF) than an affine transformation. Splines in various forms are used, and thin-plate splines are probably the most popular. Thin-plate splines have been used for surface-, landmark- and voxel based image registration [183], but they have a disadvantage. Since the thin-plate splines' radial basis functions have infinite support i.e. changing the position of control points affects not only the neighborhood of the control point in question, they also make it hard to match the local differences in detail. To get a local support the free-form deformations (FFDs) is a good choice, due to their robustness, smoothness, C^2-continuity, and their relatively simple composition of cubic B-splines [59].

The FFD is defined across the imaging domain. The model is specified by a grid $\boldsymbol{\Phi}$ of $p_x \times p_y \times p_z$ control points $\boldsymbol{\phi_{i,j,k}}$ with spacing (δ_x, δ_y, δ_z) distributed equidistantly across the domain. For each of the control points a cubic basis function with local support is positioned. The image can than be deformed by manipulating the weights $\boldsymbol{\phi_{i,j,k}}$ associated with the given basis function. Mathematically the FFDs can be written as the tensor product of 1D cubic

B-splines (here written for u, but it is silimar for v and w):

$$T_{\text{local}}(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\phi_{i+l,\,j+m,\,k+n}, \tag{6.11}$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $k = \lfloor z/n_z \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$ and $w = z/n_z - \lfloor z/n_z \rfloor$. B_0 through B_3 represent the basis functions of the B-spline:

$$B_0(u) = (1-u)^3/6, \tag{6.12}$$
$$B_1(u) = (3u^3 - 6u^2 + 4)/6, \tag{6.13}$$
$$B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6, \tag{6.14}$$
$$B_3(u) = u^3/6. \tag{6.15}$$

The control points $\mathbf{\Phi}$ control the parameters for the B-spline FFD, and the degree of nonrigid deformation that can be modeled essentially depends on the control points. Thus, a large spacing of control point can model large scale nonrigid deformations, while smaller spacing can model highly local nonrigid deformations. Consequently, the computational complexity depends on the resolution of the control point mesh i.e. the degrees of freedom. The best compromise between the degree of nonrigid deformation and computational complexity is obtained through a multilevel approach [163, 170].

6.2.2.3 Regularization

The multilevel approach is hierarchically built i.e. the resolution of the B-spline control point mesh is increased along with the image resolution in a coarse to fine way. Another advantage of the multilevel approach is that is helps avoiding local minimas. Let $\mathbf{\Phi}^1, ..., \mathbf{\Phi}^H$ denote a sequence of control point meshes with arbitrarily increasing resolution. Each control point mesh $\mathbf{\Phi}^h$ and the associated B-spline FFD define a local transformation T^h_{local} at each resolution level, and their sum constitutes the local transformation T_{local}

$$T_{local}(x,y,z) = \sum_{h=1}^{H} T^h_{local}(x,y,z). \tag{6.16}$$

To avoid calculating several B-splines, the local transformation is represented by a single B-spline FFD, whose control point mesh is progressively refined using suitable B-spline subdivision schemes [65]. A nice property of the multi-resolution B-spline FFD is that at low resolutions the single mesh FFD is not flexible enough to fold, while at high resolutions large deformations can be modeled without

folding due to the preceding lower levels. Hence, when using multi-resolution FFDs the smoothness constraint (regularizer) is no longer crucial and we can put $\lambda = 0$. Gaussian smoothing or similar of the volumes is recommended as a sort of regularization to avoid local minima.

Furthermore, non-equidistant control spacing is possible with Rueckert et. al's framework. A control point status, *active* or *passive*, can be associated with each control point at each level in the mesh hierarchy. Active control points are allowed to move during registration while passive are not. Prior knowledge from e.g. a segmentation can be used.

6.2.2.4 Similarity Measure

Many different similarity measures exists such as sum of squared differences (SSD), correlation coefficient (CC), joint entropy, mutual information (MI) and normalized mutual information (NMI) [183]. NMI is considered a robust intensity-based voxel similarity measure, and validation has shown that it works at least as well as MI and in some cases better [183, 187]. As a consequence of the general use NMI was applied in this thesis. Usually NMI is applied to images with very different intensities e.g. if a contrast agent has been used or if e.g. MRI is to be matched with CT.

The theory behind NMI is provided in the following. The marginal discrete entropy is defined as

$$H(A) = -\sum_{a \in A} p(a) \log p(a), \tag{6.17}$$

where a is the voxel intensities of image A, $p(a)$ is the marginal probability i.e. the probability of finding the intensity a in A. The joint entropy is defined as the overlap between the images A and B with voxel-intensities a and b, respectively.

$$H(A, B) = -\sum_{a \in A, b \in B} p(a, b) \log p(a, b), \tag{6.18}$$

where $p(a, b)$ is the estimated joint probability or the probability of finding a particular pair of voxel-intensities.

Joint entropy is, however, not sufficiently robust for many voxel-similarity based registrations due to the joint probability which is only defined for the region of overlap between the images. This means, that if the background or noise is a very large part of the overlapping region, the resulting registration might be

incorrect. Hence, the joint entropy usually requires a modification into e.g. MI or NMI to become a useful measure.

MI incorporates the marginal entropy in the overlapping region [200, 40].

$$MI(A,B) = H(A) + H(B) - H(A,B) \tag{6.19}$$

However, this does not fully solve the problem. Changes in the marginal entropy, e.g. by translation and rotation might affect the similarity measure. Thus, by dividing MI with the joint entropy, MI is normalized and overlap invariance is provided [187]. In terms of image registration this means that the NMI measure takes into account the changes in the intensity histograms constructed from of images A and B,

$$C_{similarity}(A,B) = NMI(A,B) = \frac{H(A) + H(B)}{H(A,B)}. \tag{6.20}$$

See also Maes et. al [119].

6.2.2.5 Optimization

The optimal transformation is found by minimizing a function associated by the parameters from the global, Θ, and the local, Φ, transformation models.

$$C(\Theta,\Phi) = C_{\text{similarity}}(B, T(A)) + \lambda C_{\text{smooth}}(T). \tag{6.21}$$

, The (dis)similarity measure $C_{\text{similarity}}$ represents the cost associated with the voxel similarity between image pairs. The smoothness term C_{smooth} or regularizer is defined as the bending energy of a thin-plate spline weighted by a constant λ. The weighting parameter λ defines the trade-off between the alignment of the two volumes and the smoothness of the transformation. A simple gradient descent technique is used in at each stage to find the local optimum [163].

6.2.2.6 Collecting the Pieces

For computational efficiency, the global model parameters Θ are iteratively optimized first based on the similarity measure. During the subsequent stage, the nonrigid transformation parameters Φ are optimized as a function of the cost function in Equation 6.21. A simple gradient descent technique is used in at each stage to find the local optimum [163]. The resolution of the control point mesh is the limiting factor for the performance of the method. Global transformations can be modeled using a coarser grid and still perform well, while that is not the case for local transformations, which often require a finer grid.

6.2.2.7 Resampling

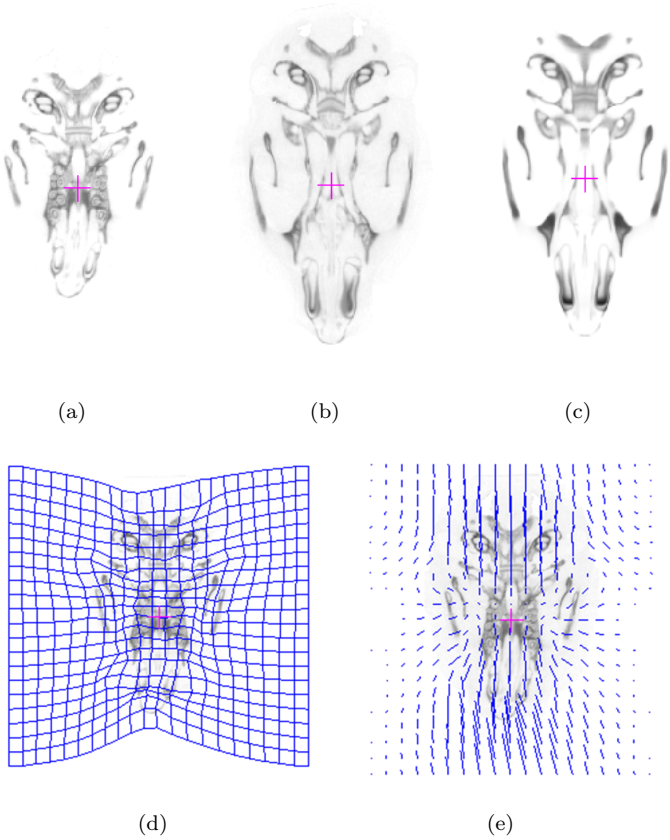


Figure 6.1: Example of a FFD-based non-rigid registration. A 3D CT image of a 4-week Crouzon mouse (source) is registered to a 3D CT image of a normal mouse (target). Single-slice sagittal views are presented. (a) Source image, (b) source image deformed to resemble the target image (interpolation method: nearest neighbor), (c) target image. (d) Resulting deformation grid and (e) corresponding deformation vectors at control point positions.

When the transformation is estimated, the output image is constructed, i.e. every voxel in the output image is assigned a value. This process is called resampling because we sample the values from the source into the target voxel locations. While voxels are distributed in a regular, discrete lattice, the typical transformation model has a continuous range. Normally, the transformed row and column numbers will be non-integers, therefore it is necessary to interpolate between the voxels to find the voxel value assigned to the output image [67]. Different

interpolation schemes exists, the most commonly used is linear interpolation. For binary images such as masks nearest neighborhood is usually preferred. A concrete example of a transformation is given in Figure 6.1.

6.3 Statistical Methods

An important issue when analyzing multiple findings in image analysis is to test the findings for statistical significance. This is a non-trivial task as voxel correspondences are highly correlated and multiple comparisons need to be corrected for. In general, the problem is to simultaneously test several correlated hypotheses. Two methods were applied in this thesis: Firstly, local False Discovery Rate (local FDR); a method for large-scale simultaneous hypothesis-testing problems, which occur when testing for statistical significance in multiple findings. The second approach was a Multivariate Analysis of Variance (MANOVA). A MANOVA is used in cases where there is more than one dependent variable, and where the dependent variables cannot simply be combined. The local FDR will be described in this section including some examples; the MANOVA has been described in detail in Section 12.4.1.

6.3.1 Local False Discovery Rate

Large-scale simultaneous hypothesis testing problems occur when testing for statistical significance in multiple correlated findings from e.g. image analysis algorithms. Since there is often a vast amount of findings, it is an important issue to correct for the correlations. Efron [57] addressed the problem of simultaneously testing several correlated hypotheses in the analysis of DNA microarrays with the use of local FDR. FDR theory relies on p-values i.e. on null-hypothesis tail areas. Thus, local FDR estimates the expected percentage of false predictions in a set of predictions. According to Efron, huge datasets make it possible to replace a theoretical null-hypothesis by estimating a null-hypothesis from the data itself.

The methodology requires the estimation of a statistical distribution for the input data and corresponding p-values at each data point. For every i^{th} data point, e.g. surface vertex, a p-value p_i is estimated and the corresponding test statistic z-value is estimated by

$$z_i = \Phi^{-1}(p_i),\ i = 1, ..., m, \tag{6.22}$$

where Φ is the cumulative distribution function (cdf). Assuming the data

is Gaussian distributed, a promising z-value histogram would have a normal-shape peak presumably holding the majority of uninteresting 'null' vertices (for instance non-growth areas), while long tails hold the interesting 'non-null' vertices. With the assumption of dividing the z-values into the classes 'interesting' and 'uninteresting', z_i has density $f_0(z)$ or $f_1(z)$ with prior probabilities p_0 and $p_1 = 1 - p_0$ depending on the class. The mixture density $f(z)$ is then estimated by fitting a smooth curve or spline to the z-histogram

$$f(z) = p_0 f_0(z) + p_1 f_1(z). \tag{6.23}$$

The posterior probability, that a case z_i given z, falls in the uninteresting class follows from Bayes theorem

$$fdr(z) \equiv \Pr\{\text{uninteresting}|z\} = \frac{p_0 f_0(z)}{f(z)}. \tag{6.24}$$

Efron then defines the local false discovery rate as

$$fdr(z) \equiv \frac{f_0(z)}{f(z)}, \tag{6.25}$$

where p_0 in Equation 6.24 is ignored, because Efron assumes p_0 is close to 1. He further states that the best results are achieved for $p_0 \geq 0.90$. Equation 6.24 provides an upper bound on $\Pr\{\text{uninteresting}|z\}$. Moreover, the empirical null-hypothesis, obtained by estimating the mean and standard deviation from $f(z)$, can replace the theoretical null-hypothesis. But Efron urges caution since the use of the theoretical null-hypothesis in some cases leads to findings whereas the empirical does not – situations in the reverse direction might also occur.

6.3.1.1 Examples

To give an example of the use of local FDR, we compare point-wise deformations from the surface of a 6-week-old normal atlas[1] to all the other mice within the specific group (Crouzon or normal, read Chapter 12). The analysis is based on the magnitude of the displacement vectors, and it is assumed that each component of the three-dimensional deformation vector follows a normal distribution.

For each mouse-type (Crouzon or normal) the mean magnitude of the growth vectors (GV) of group G_1 (4-weeks) was compared to that of group G_2 (6-weeks). The hypothesis:

$$\begin{aligned} H_{0,i} &: \bar{GV}_i^{G_1} = \bar{GV}_i^{G_2},\ i = 1, ..., m \\ H_{1,i} &: \bar{GV}_i^{G_1} \neq \bar{GV}_i^{G_2},\ i = 1, ..., m \end{aligned} \tag{6.26}$$

[1] An atlas represents the average anatomy.

where

$$\bar{GV}_i^{G_1} = \frac{1}{N_{G_1}} \sum_{j \in G_1} GV_i^j \tag{6.27}$$

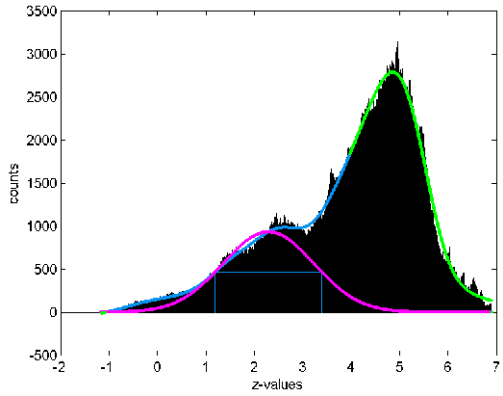
is the mean magnitude of the growth vectors in point i across mouse j in group G_1. N_{G_1} is the number of mice in group G_1, and m is the number of vertices ($m = 945.284$ in this case). $\bar{GV}_i^{G_2}$ can be written in a similar way.

To estimate the probability distribution of the growth values, the displacement vector components were assumed to be normally distributed. Hence, the squared magnitude of the growth vector approximately follows a χ^2-distribution with 3 degrees of freedom. The significance of the difference(s) between the groups can now be estimated by an F-test, since the ratio of two χ^2-distributions follows an F-distribution. Given the squared magnitude of the growth vector $\bar{GV}_i$, the F-statistics becomes,

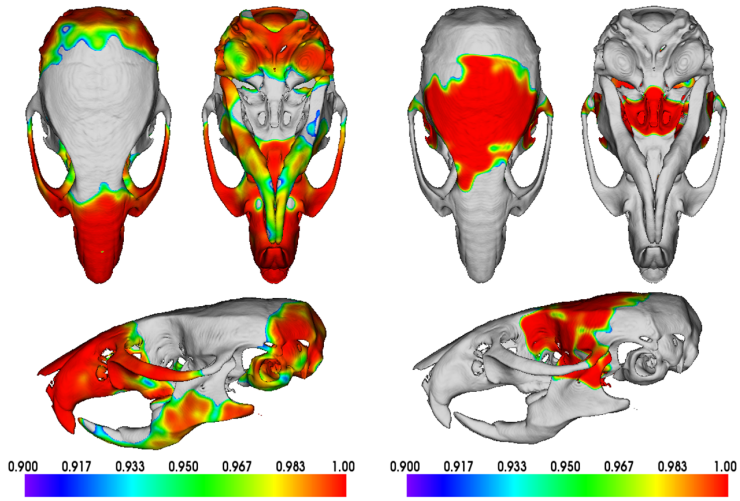
$$\frac{\frac{|\bar{GV}_i^{G_1}|}{2N_{G_1}}}{\frac{|\bar{GV}_i^{G_2}|}{2N_{G_2}}} \sim F(2N_{G_1}, 2N_{G_2}),\ i = 1, ..., m, \tag{6.28}$$

where $\bar{GV}_i^{G_1}$ is given by Equation 6.27. p-values in each point can now be obtained by evaluating the F-statistics.

As stated previously, a promising z-value histogram will have a normal-shape peak presumably containing the majority of not-interesting "null"-vertices (non-growth or steady-state areas), while long tails hold the interesting "non-null"-vertices (growth). With this knowledge it is now possible to interpret the outcome from the local FDR analysis, see Chapter 12. Figure 6.2(b) reveals the significant areas of change (growth) between 4 and 6-week normal mice. In Figure 6.2(c) the significant areas with no growth are shown. The Crouzon group was interpreted in a similar way, see Figure 6.3.



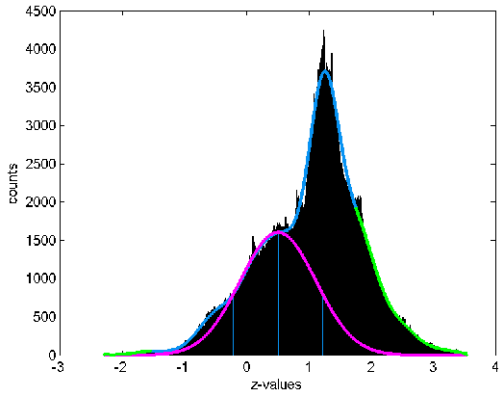
(a) z-value histogram with the empirical null-hypothesis fitted to the left peak.



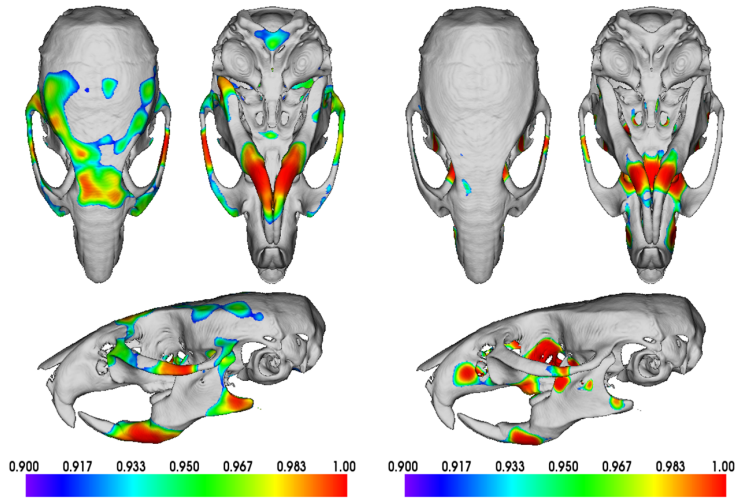
(b) Significance determined from left peak in (a).

(c) Significance determined from right peak in (a).

Figure 6.2: Local FDR analysis on wild-type mice. A group-wise comparison of deformation vector magnitudes in each point of 4 and 6-week wild-type mice. (a) Example of a z-value histogram fitted to the left peak. The blue curve is the fitting spline, $f(z)$; the magenta curve represents the empirical null-hypothesis, $f_0(z)$; and the green curve denotes the observations with $fdr(z) \leq 0.10$. (b) 1-fdr displayed on the atlas surface. The positions where 1-$fdr \geq 0.9$ correspond to the bins under the green curve in (a). (c) shows the significant areas from the z-value histogram with the empirical null-hypothesis fitted to the right peak. In (b) and (c) no significance is gray, little significance purple and large significance red.



(a) z-value histogram with the empirical null-hypothesis fitted to the left peak.



(b) Significance determined from left peak in (a).

(c) Significance determined from right peak in (a).

Figure 6.3: Local FDR analysis on Crouzon mice. A group-wise comparison of deformation vector magnitudes in each point of 4 and 6-week with mirrored Crouzon mice. (a) Example of a z-value histogram fitted to the left peak. The blue curve is the fitting spline, $f(z)$; the magenta curve represents the empirical null-hypothesis, $f_0(z)$; and the green curve denotes the observations with $fdr(z) \leq 0.10$. (b) 1-fdr displayed on the atlas surface. The positions where $1 - fdr \geq 0.9$ correspond to the bins under the green curve in (a). (c) shows the significant areas from the z-value histogram with the empirical null-hypothesis fitted to the right peak. In (b) and (c) no significance is gray, little significance purple and large significance red.

6.3.2 Multivariate Analysis of Variance

Given that the data are mutually independent, identically distributed and Gaussian, a linear model can be set up for modeling growth. It is assumed that every observation $\bar{\boldsymbol{X}}_{ijv}$ can be written as

$$\bar{\boldsymbol{X}}_{ijv} = \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{a}}_i + \bar{\boldsymbol{b}}_j + \bar{\boldsymbol{ab}}_{ij} + \bar{\boldsymbol{\epsilon}}_{v(ij)}, \tag{6.29}$$
$$\sum \bar{\boldsymbol{a}}_i = 0, \ \sum \bar{\boldsymbol{b}}_j = 0, \ \sum_i \bar{\boldsymbol{ab}}_{ij} = \sum_j \bar{\boldsymbol{ab}}_{ij} = 0,$$

where $\bar{\boldsymbol{a}}_{\boldsymbol{i}}$ denotes the age, $\bar{\boldsymbol{b}}_{\boldsymbol{j}}$ the mouse-type and $\bar{\boldsymbol{ab}}_{\boldsymbol{ij}}$ the interaction between age and mouse-type. Table 12.2 shows the scheme for carrying out the MANOVA.

The Mahalanobis distance was used to test for equal means. The interaction was tested with a Hotelling's T^2. For more details on the MANOVA, read Section 12.4 in paper 12.

6.4 Mesh Generation using Delaunay Triangulation

Mesh generation is a huge research field and there are many subjects related to this matter [30]. Triangulation in three dimensions is called tetrahedralization. A tetrahedralization is a partition of the input domain, point set or polyhedron, into a collection of tetrahedra, that meet only at shared faces (points, edges, or triangles) [18]. Tetrahedralization turns out to be significantly more complicated than triangulation, and as a consequence only 2D is considered in here. Moreover, the focus will be on Delaunay triangulation as this triangulation has some nice properties e.g. Min Max Angle (MMA). MMA is a common problem in triangulations which means that some triangles become too skinny i.e. one angle becomes very small and makes the triangle long and narrow. As skinny triangles can cause numerical instability, such triangles are best avoided; here the Delaunay triangulation is a good choice.

In 2D, a triangulation usually starts with a set of points, $p_i \in \mathcal{P}$, which shall be connected by edges, $e_i \in \mathcal{E}$, such that these edges form triangles, $t_i \in \mathcal{T}$. The goal of the triangulation process is then to form a unified entity from the given points p_i.

The Delaunay triangulation divides a point set on a plane into triangles such that no other point is inside a circle defined by the three vertices of the triangle. The

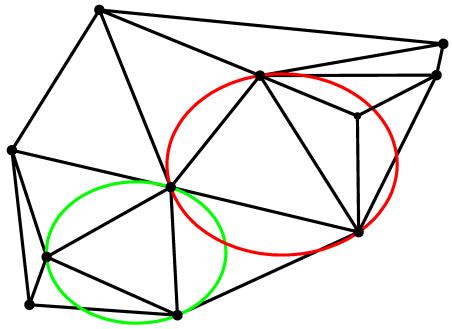


Figure 6.4: The triangle within the green circle satisfies the Delaunay criterion, while red does not as there are four points within the circle.

Delaunay criterion states that the circumcircle (the circle defined by the three vertices in a triangle) of a triangle may not contain other points, but the three belonging to the triangle itself. Other points are accepted on the perimeter of the circle, but not inside it. If there are more than three points on the perimeter of the circumcircle, the Delaunay triangulation is not unique. This is the case, for example, when triangulating the four points of a rectangle. The Delaunay criterion must hold for all triangles in the triangulation if the triangulation is a Delaunay triangulation. Moreover, it is possible to construct a Delaunay triangulation of any set of points, provided that the set is not degenerate i.e. all the points may not lie in the same place or on the same line. Various algorithms exists for Delaunay mesh generation, see [30]. An example of a triangulation is given in Figure 6.4.

According to [18] any polyhedron can be triangulated, however, Delaunay triangulation in its pure form can only be extended to 3D if the polyhedron equivalent to a sphere (convex). The concept of circumcircles are extended to circumspheres. Finding the two dimensional Delaunay triangulation of a set of point is equivalent of projecting the set on a 3D paraboloid ($z = x^2 + y^2$) and finding the convex hull of that point set. This concept can also be generalized to more dimensions.

6.5 Mixture of Gaussians

Sometimes a single Gaussian distribution is not enough to model the data, whereas a superposition of two or more Gaussians gives a better characterization.

Recall that the Gaussian distribution is defined as

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \tag{6.30}$$

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{2\pi^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \tag{6.31}$$

where x, μ, and σ^2 defines a single variable, the mean, and variance, respectively, in 1D. In the multivariate case, $\boldsymbol{x}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ defines a D-dimensional vector, a D-dimensional mean, and a a $D \times D$ covariance matrix, respectively. $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$. Let a superposition of K Gaussian densities be of the form

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{6.32}$$

where π_k are the mixing coefficients, and each Gaussian density $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ has its own mean $\boldsymbol{\mu}_k$ and variance $\boldsymbol{\Sigma}_k$.Equation 6.32 is called a mixture of Gaussians.

Introduce a K-dimensional binary random variable $\boldsymbol{z}$ with a 1-of-K representation, in which a particular element $z_k = 1$ and all other elements are 0. Thus, the values of $\boldsymbol{z}$ satisfy $z_k \in \{0, 1\}$ and $\sum_k z_k = 1$ leaving K possible states for $\boldsymbol{z}$ according to which element is nonzero. Defining the joint distribution $p(\boldsymbol{x}, \boldsymbol{z})$ in term of a marginal distribution $p(\boldsymbol{z})$ and a conditional distribution $p(\boldsymbol{x}|\boldsymbol{z})$. The marginal distribution is specified in terms of the mixing coefficients π_k

$$p(z_k = 1) = \pi_k, \{\pi_k\} : 0 \leq \pi_k \leq 1 \text{ and } \sum_{k=1}^{K} \pi_k = 1, \tag{6.33}$$

where the conditions on π_k is necessary in order to be valid probabilities. Because $\boldsymbol{z}$ uses a 1-of-K representation, the distribution in (6.33) can be written as

$$p(\boldsymbol{z}) = \prod_{k=1}^{K} \pi_k^{z_k}. \tag{6.34}$$

Similarly, the conditional distribution of $\boldsymbol{x}$ given a particular value for $\boldsymbol{z}$ is a Gaussian

$$p(\boldsymbol{x}|z_k = 1) = \mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \tag{6.35}$$

which can be written as

$$p(\boldsymbol{x}|\boldsymbol{z}) = \prod_{k=1}^{K} \mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)^{z_k}. \tag{6.36}$$

The joint distribution is given by $p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z})$, and the marginal distribution of $\boldsymbol{x}$ is obtained by summing the joint distribution over all states of $\boldsymbol{z}$, which give

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{z}} p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \tag{6.37}$$

where (6.34) and (6.36) has been used. Thus, the marginal distribution of $\boldsymbol{x}$ is a Gaussian mixture of form (6.32). It might seem that nothing has been gained, finding an equivalent formulation of the Gaussian mixture involving a explicit hidden variable $\boldsymbol{z}$. However, now the joint distribution can replace the marginal distribution distribution $p(\boldsymbol{x})$ and significantly simplify the expectation-maximization (EM) algorithm [21].

Moreover, the conditional probability of $\boldsymbol{z}$ given $\boldsymbol{x}$ is important. Let $\gamma(z_k)$ denote $p(z_k = 1|\boldsymbol{x})$, then using Bayes' Rule

$$\begin{aligned} \gamma(z_k) \propto p(z_k = 1|\boldsymbol{x}) &= \frac{p(z_k = 1)p(\boldsymbol{x}|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\boldsymbol{x}|z_j = 1)} \\ &= \frac{\pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \end{aligned} \tag{6.38}$$

Once $\boldsymbol{x}$ has been observed, π_k can be seen as the prior probability of $z_k = 1$, and $\gamma(z_k)$ as the corresponding posterior probability [21].

Suppose the independent, identically distributed data $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$ is a set of observations, and we wish to model this data using a mixture of Gaussians. Then $\boldsymbol{X}$ is a $N \times D$ matrix in which the n^{th} row is given by $\boldsymbol{x}_n^T$. Similar, the corresponding hidden variables will be a $N \times K$ matrix $\boldsymbol{Z}$ with rows $\boldsymbol{z}_n^T$. From (6.32) the log of the likelihood function is given by

$$\ln p(\boldsymbol{X}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum^{N} n = 1 \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}. \tag{6.39}$$

Certain significant problems are associated with the maximum likelihood framework when applied to Gaussian mixture models. For instance singularities pose a problem, see [21] for more details.

The EM algorithm seeks to maximize the likelihood function with respect to the parameters (means, covariances and mixing coefficients). For Gaussian mixtures the EM algorithm can be summarized as follows (more details are given in [21]):

1. Initialize the means $\boldsymbol{\mu}_k$, covariances $\boldsymbol{\Sigma}_k$ and mixing coefficients π_k, and evaluate the initial value of the log-likelihood function.

2. **E step** Evaluate the responsibilities $\gamma(z_{nk})$ using the current parameter values,

$$\gamma(z_n k) = \frac{\pi_k \mathcal{N}(\boldsymbol{x_n}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x_n}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

3. **M step** Re-estimate the parameters using the current responsibilities

$$\boldsymbol{\mu}_k^{new} = \tfrac{1}{N_k} \sum n = 1^N \gamma(z_n k) \boldsymbol{x}_n$$

$$\boldsymbol{\Sigma}_k^{new} = \tfrac{1}{N_k} \sum n = 1^N \gamma(z_n k) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{new}\right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{new}\right)^T$$

$$\pi_k^{new} = \tfrac{N_k}{N},$$

where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$.

4. Evaluate the log-likelihood

$$\ln p(\boldsymbol{X}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum^N n = 1 \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\},$$

and check for convergence of either the parameters or the log-likelihood. If the convergence criterion is not satisfied return to step 2.

6.6 Image Segmentation using Graph Cuts

Image segmentation is defined as separating the image into specific regions, meaningful for a specific task. Image quality and acquisition artifacts directly affect the segmentation quality and success.
Different methods and techniques exist for segmentation and classification of images. This is an extensive subject and we only touch some of the methods briefly. Overall, the methods can be separated into techniques based on: 1) analysis of gray-level histograms (e.g. thresholding); 2) edge-based and region-based methods (e.g. dynamic programming, border tracing, region growing, connected components labeling); 3) classification methods (e.g. Bayesian classification, Gaussian Mixture Models, clustering), and finally, 4) image segmentation using deformable models (e.g level set method, active shape models, atlas-based segmentation, graph cuts). [184] is recommended for more detailed information and references on the different topics. Pixels, voxels or resampled versions of these will be termed points.

In this short introduction to image segmentation, we will briefly touch upon thresholding, Bayesian classification, and Markov Random Fields (MRF). Thresholding is probably the simplest segmentation method, separating the image into,

for instance object and background, see also Section 6.1. Thresholds can be determined manually from the intensity histogram or automatically. A well-known example of an automatic method is Otsu's method [136].

Classical Bayesian classification, or more formally Bayesian Maximum Likelihood Classification, is based on Bayes formula, Equation 6.40, and is a *supervised, parametric* method. The formula enables us to calculate the *posterior probability*, $P(x_i|v)$, i.e. the probability of a given pixel belongs class x_i given intensity v.

$$P(x_i|v) = \frac{P(v|x_i)P(x_i)}{P(v)}, \tag{6.40}$$

where x_i denotes a class, and v the intensity of a given voxel. $P(x_i)$ denotes the *prior probability* of the classes. $P(v|x_i)$ is the *class conditional probability*. This is the probability of a given pixel having intensity v given that the pixel belongs class x_i. Finally, $P(v)$ acts as a normalizing constant and is set to one. In practice, $P(x_i|v)$, can be calculated as

$$\begin{aligned} P(x_i|v) &= P(x_i)\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{(v-\mu_i)^2}{2\sigma_i^2}\right) \\ &= \ln P(x_i) - \ln\sigma_i - \frac{(v-\mu_i)^2}{2\sigma_i^2}, \end{aligned} \tag{6.41}$$

assuming the data are Gaussian distributed. μ_i and σ_i are the pixel value average and the standard deviation of class x_i. Thus, by adding classes prior information is including in the segmentation compared to thresholding [138].

Over the past years graph-based approaches have been playing an important role in image processing and a wide variety of computer vision problems such as image restoration [27, 25, 73, 86], stereo and motion [20, 27, 25, 87, 95, 100, 159, 160] and medical imaging [28, 24, 96, 94]. In the following we will focus on the binary labeling problem of the Potts model (i.e. the Ising model) regarding image segmentation. Typical graph algorithms that are exploited for image segmentation include minimum spanning trees; shortest paths; and graph cuts [112].

6.6.1 Markov Random Fields

Markov Random Fields (MRF) is a probabilistic framework that can be used for modeling an image including context. Clique potential functions specify the probability of assigning a particular classification to a voxel given the intensities and the classification of its neighbor. Compared to the the classical Bayesian

classification, MRF add neighbors to the segmentation problem [184]. MRFs are very general can can model a range of phenomena, e.g. image segmentation, boundary finding, image restoration [70]. Formulating MRFs in a Bayesian framework consists of four stages [19]:

1. Construction of a prior probability distribution $P(x_i)$.

2. Formulation of a model $P(v|x_i)$ that describes the distribution of the observed data given any particular realization of the prior distribution.

3. Combination of the prior and the tissue model into the posterior distribution by Bayes theorem, Equation 6.40.

4. Drawing inference based on the posterior distribution.

The aim is to find the function that maximizes the posterior probability $P(x_i|v)$.

Given a graph of m points a MRF consists of three sets ; a set $S = \{s_t\}_{t=1}^{m}$ of sites (points,pixels), a neighborhood system N and a set of random variables X [99, 27, 35]. The neighborhood system $N = \{N_s|s \in S\}$, where each N_s is a subset of sites of S which form the neighborhood of site s. The random field $X = \{X_s|s \in S\}$ consists of random variables X_s that take on a value x_s from a set of labels $L = \{l_h\}_{h=1}^{k}$. A joint event $\{X_1 = x_1, .., X_m = x_m\}$ is abbreviated as $X = x$, where $x = \{x_s, s \in S\}$ is a configuration of X,corresponding to the realization of the field. Denote $P(X = x)$ as $P(x)$ and $P(X_s = x_s)$ as $P(x_s.)$. In order to be a MRF, X must satisfy the Markov property:

$$P(x_s|x_{S-s}) = P(x_s|x_{N_s}), \forall s \in S, \tag{6.42}$$

where $S - s$ is the set of sites without site s. The second requirement for a MRF is that all configurations are possible, i.e. $P(x) > 0, \forall x \in X$. This means that the state of each random variable depends only on the state of its neighbors. The fact that a particular point label depends on the labels of its neighbors allows modeling the optimization problem as a MRF. The key result concerning MRFs is the Hammersley-Clifford theorem. This states that the probability of a particular configuration $Pr(x) \propto exp(-\sum_C V_C(x))$, where the sum is over all cliques in the neighborhood system N. Here, V_C is a clique potential, which describes the prior probability of a particular realization of the elements of the clique C. A clique is a subset of S for which every pair of sites are neighbors. We will restrict our attention to MRFs whose clique potentials involve pairs of neighboring pixels,

$$Pr(x) \propto \exp\left(-\sum_{s \in S}\sum_{j \in N_s} V_{s,j}(x_s, x_j)\right). \tag{6.43}$$

In general, the field X is not directly observable in the experiment. We have to estimate its realized configuration x based on the observation D, which is related to x by means of the likelihood function $Pr(D|x)$. D is here the joint event $I_s = v_s$ over all $s \in S$, where I_s denotes the observable intensity at pixel s and v_s is its particular realization. If X_s denotes the true intensity at s then

$$P(D, x) = \Phi(v_s, x_s) = \prod s \in Sg(v_s, x, s), \tag{6.44}$$

where $g(v_s, x, s) = p(I_s = v_s, X_s = x_s)$ represent the noise model, which returns high values for low probabilities and vice versa ($0 \leq g \leq 1$).

We wish to obtain the configuration $x \in L$ that maximizes the posterior probability $P(x|D)$. Bayes' Rule tells us that $P(x|D) \propto Pr(D|x)Pr(x)$. It follows that our MAP estimate f should minimize the posterior energy function

$$E(x) = \sum_{s \in S} \sum_{j \in N_s} V_{s,j}(x_s, x_j) - \sum_{s \in S} \ln(g(v_s, x_s)). \tag{6.45}$$

Graph cuts will be used for finding an optimal solution to the MRF [27, 70].

6.6.2 Graphs in Image Segmentation

In 1989 Greig et al. [73] showed that min-cut/max-flow algorithms from combinatorial optimization can be used to solve the classical pixel-labeling problem. The input is a set of sites (points) S, and a set of labels L (binary, in our case). The goal is to find a labeling, x, mapping S to L, such that x minimizes some energy function. The standard energy form can be represented as a posterior energy in a MAP-MRF framework[2]

$$E(x_s) = E_{data}(x_s) + E_{smooth}(x_s) = \sum_{s \in S} D_s(x_s) + \sum_{s,j \in N_s} V_{s,j}(x_s, x_j), \tag{6.46}$$

where D_s is the data penalty function, that measures the cost of assigning the label x_s to the point s; $V_{s,j}$ is the smoothness penalty and measures the cost of assigning the labels x_s, x_j to the adjacent points s,j in N_s, [101, 26, 70].

The general approach is based on the formation of a weighted graph, $G = (V, E)$ with *vertex set* V and *edge set* E defining the neighbors. The vertices $v \in V$ usually correspond to pixels, voxels or resampled versions of these. All edges, connecting the vertices in a neighborhood system, are assigned a non-negative weight or cost that specify the measure of preference of the connections [112].

[2]MAP-MRF stands for Maximum *A Posterior* estimation of a Markov Random Field.

In the context image segmentation the terminal nodes, the *source* s and the *sink* t, usually represents labels e.g. object and background that can be assigned to voxels. The edge set E consists of two types of edges: *n-links* and *t-links* [112]. *n-links* connect pairs of neighboring points whose cost is derived from the smoothness term V in Equation 6.46 i.e. there is a penalty for discontinuity between the points. *t-links* connect points with terminal nodes (the labels), and the cost is associated with assigning the corresponding label to a point. The *t-link* cost is usually derived from the data term D_p in Equation 6.46.

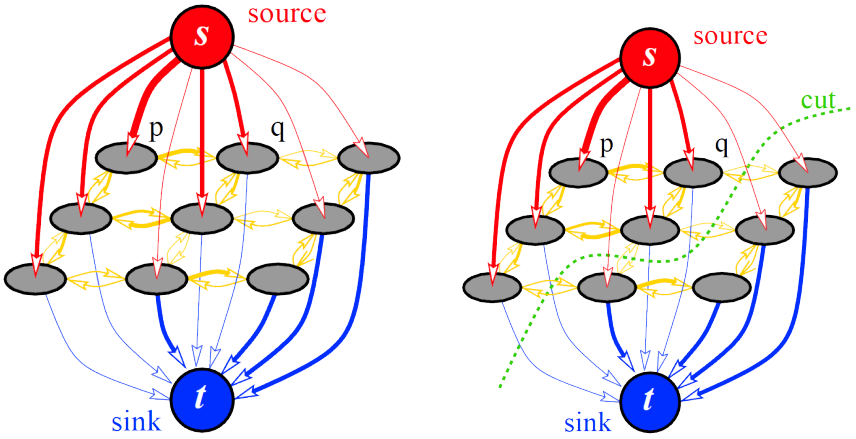
The definition of a graph cut is a separation of the set of vertices into two disjoint subsets. Such a cut is also called a *s-t-cut* or *source-sink separation*. All graph edges have an associated cost: In the case of directed graphs the edge (i, j) has a different cost associated than the reverse edge (i, j); for undirected graphs the costs are similar. The *s-t-cut* thus partitions the vertices of the graph such that no path can be established between the *source* and the *sink*. Furthermore, the cost of a cut is equal to the cost of the edges in the cut [26, 112, 101].

6.6.3 Min-cut/Max Flow Problems

Various algorithms exist for min-cut/max-flow problems with two terminals. These algorithms can be divided into two groups: Goldberg-Tarjan-style: 'push-relabel', or the Fold-Fulkerson style: 'augmenting paths' [26]. Boykov and Kolmogorov [26] present an analysis of different min-cut/max-flow algorithms used for image analysis and computer vision problems, proving that their algorithm performs faster than others in most of the tests performed. This augmenting paths algorithm was used in this project to find the global optimum of the Ising model, while the Potts model (multi-labels) usually are solved via α-expansion.

Standard 'augmenting paths' based algorithms work by pushing flow from the source to the sink through non-saturated paths in the graph. A path is considered to be saturated when no flow can be sent from source to sink through that path, i.e. if at least one of the edges is saturated. Loosely speaking maximum flow can be regarded as the amount of water which can be pushed from the source to the sink through water pipes, where the graph edges are interpreted as pipes with capacities equal to the cost associated with the edge weights. An intuitive example is given in Figure 6.5; here a min-cut/max-flow is illustrated on a graph.

In the Boykov-Kolmogorov approach two search trees are constructed - one starting from the source and one starting from the sink. These trees are reused compared to building them from scratch when searching paths of length k+1. The disadvantage of this approach is that the found augmenting paths are not



(a) A 2D graph G.

(b) A cut on G.

Figure 6.5: Example of a two-label graph in an Ising model. Edge costs are reflected by the arrows' thickness. p and q are neighboring points in a 3×3 image. (6.5(b)) show the *s-t-cut* that partitions the *source* from the *sink*. Adopted from [26].

necessarily the shortest ones. The Boykov-Kolmogorov approach performs better in computer vision problems due to the reuse of the search trees.

The information of the flow among the edge of the graph G is kept using a residual graph. The residual graph has exactly the same topology as the graph G, but the capacity of the edge is represented by the residual capacity given the flow amount run through that edge. In each iteration the shortest path from source to sink through non-saturated edges of the residual graph is chosen. When a path is found the maximum possible flow, which saturates at least one edge on the path is pushed through the path and the residual capacities of the edges on that path are reduced by the pushed flow and the residual capacities of the inverse edges are increased by the flow. The maximum flow is computed by adding the pushed flow value to a total flow value. The algorithm stops when any path between source and sink has at least one edge that is saturated. An edge is saturated when the weight associated to that edge in the residual graph is equal to 0 (no flow value can be sent through that edge). The minimum cut is found by choosing the saturated edges with the smallest capacity on all the paths from source to sink.

The two search trees, with roots at the source and at the sink, are denoted by

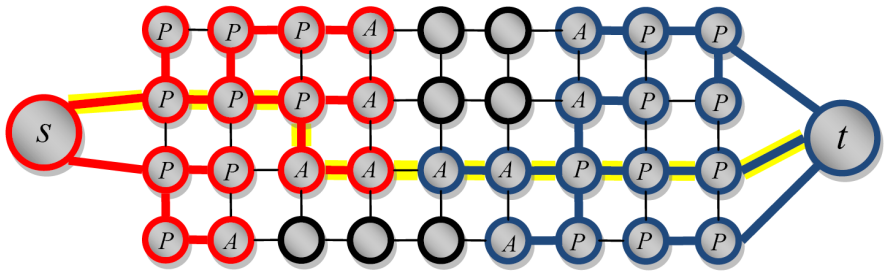


Figure 6.6: Search tree setup for a short overview of the min-cut/max-flow algorithm and its terminology. Red nodes, S, are connected to the search tree with root in the source, whereas blue nodes, T, belong to the sink. The labeling with A and P specifies the active and passive nodes, respectively. The free nodes are marked with black. The found path from the source, s, to the sink, t, at the end of a growth stage is marked with yellow. uAdopted from [26].

S and T. In tree S all the edges from each parent node to its children nodes are non-saturated. In tree T all the edges from the children nodes to its parent nodes are non-saturated. The nodes that do not belong to either of the trees are called "free" nodes and in the beginning of the algorithm the intersection between the two search trees is the empty set, tree S containing only the source node and tree T containing only the sink node. The nodes in trees T and S can be denoted either "active" or "passive". Active nodes are allowed to grow by adding children nodes on non-saturated edges. Also active nodes can also come in contact with nodes from the other tree, signaling that an augmenting path has been found. In Figure 6.6 will provide a short overview of the algorithm and its terminology. The algorithm consists in repeating in an iterative manner the three steps described below:

1. Growth In the growth step the trees S and T are expanded. The expansion process is done through active nodes exploring adjacent non-saturated edges and acquiring new children from the set of the free nodes. The acquired nodes become active nodes in the corresponding search tree. An active node becomes passive when all the neighbors of a given active node are explored. The growth step stops when an active node finds a neighboring node belonging to the other tree.

2. Augmentation After finding an augmentation path in the growth step, the path is augmented by pushing the largest flow possible (the minimum capacity of the edges on the path). As some of the edges will become saturated, some of

the nodes in the S and T trees may become orphans. A node is orphan if the edges linking these nodes to its parents are saturated. During this step the S and T trees may be split into forests where the orphan nodes behave as roots of the trees in the forest.

3. Adoption After the augmentation step, the S and T trees might be split into forests. In the adoption step, the orphan nodes, that act as roots for trees in the formed forests will either be adopted to the S or T trees or declared free nodes. In this step a new valid parent is searched for each orphan node. The parent has to belong to the same tree as the orphan and should be connected to this through a non-saturated edge. If no parent is found for the orphan node, the node is removed from the tree it belongs to (S or T), being declared a free node. In this case all its former children are also labeled as orphan nodes and need to be submitted to the adoption process. The adoption step terminates when all orphan nodes have been either adopted or labeled as free. The role of this step is to restore the structure of the S and T trees to single-tree with roots in nodes s and t.

These three steps are iterated, until none of the search trees S and T can grow and the trees are separated by saturated edges only. See [26] for more details.

Chapter 7

Discussion and Conclusion

In Section 1.1 the objectives of the present thesis were presented, namely the application of image analysis methods in order to answer clinical questions like:

1. Quantification of change related to growth and/or treatment
2. Determination of phenotypes and their variation within and between populations
3. Quantification of severity of a craniofacial malformation through comparison to a (normal) control group

This thesis, including the four articles presented in Chapters 9-12, addressed four concrete problems in craniofacial research. In each problem, a set of appropriate image analysis methodologies and tools have been identified and, when necessary, improved and developed further. Finally, the methods have been applied to solve a problem, in each case providing new knowledge in the clinical or biological field. Thus, the thesis brings novel contributions both to the technical/imaging and medical/odontological fields.

Key requirements that were specified as desirable were:

a) full automation

b) high speed (efficiency)

c) high spatial detail

d) high accuracy

Automation and high speed are important to be able to handle the large datasets typically involved, and necessary in order to make the methods usable in busy clinical environments. High spatial detail is desirable in order to take advantage of 3D information available through the use of modern scanners, and to supersede present methods using e.g. sparse sets of landmarks or 2D X-rays. Finally, high accuracy must be ascertained through validation.

Table 7.1 summarizes the four problems in craniofacial research addressed in the present thesis, stating the main image analysis methodologies used and developed, and the main application areas with major achievements in the medical/odontological domain.

The main methodologies have been image segmentation using graph cuts (Chapter 9) and non-rigid registration (Chapters 10-12).

In Chapter 9, segmentation using graph cuts was proposed for estimation of intracranial volume (ICV) in craniofacial anomalies. The ICV is an important parameter where current methods of estimation typically require tedious manual segmentation of an hour's duration or more.

In MRI fully automatic methods for ICV estimation (often called TIV; or total intracranial volume) exist [10, 11, 92, 139, 166, 180]. Most of these methods [10, 11, 92, 139] are based on segmentation techniques which are not applicable to CT, due to the low contrasts in the soft tissue. The remaining two methods, [166] and [180] are based on morphological operations in combination with graph cuts and the deformable models (Brain Extraction Tool), respectively. Returning to the only automatic method for ICV estimation in CT scans in [208], Section 5.1, thresholding with morphological operations are mostly only semi-automated; the user is normally involved in helping to choose the threshold(s) used in the initial segmentation [180]. Through trial and error [208] found that CT values between 1010 and 1250 should be set to one, and the other values to zero. A second problem of thresholding with morphological operations is that it is very hard to produce a general algorithm for the morphological step that will successfully separate brain from non-brain tissue [180]. In [208], no structuring element is given; only the number of times it is applied, which might change

Chapter	Data	Methods	Application or achievement	Objective (see list above)	Requirement (see list above)
8	Unilateral Coronal Synostosis (UCS)	Image segmentation using graph cut	Intra-Cranial Volume (ICV) quantification in UCS	2	a,b,d
9	Unilateral Cleft Lip and Palate (UCLP)	Non-rigid registration	Mapping of primary morphology in UCLP	2	a,c,d
10	Unilateral Cleft Lip and Palate (UCLP)	Non-rigid registration and knowledge base	Effect of lip adhesion in UCLP; UCLP growth	1	a,c,d
11	Mouse model of Crouzon Syndrome	Non-rigid registration	Growth pattern in Crouzon mice compared to normal	1,2,3	a,c,d

Table 7.1: Summary of the medical problems in clinical research, examined in the present thesis, listed according to Chapter.

with the resolution of the CT scan. Moreover, the validation of the method is limited. Apparently, the method was applied to five CT scans, but only one is used for validation. The relative error for the post-op skull, i.e., the one with holes has a relative error of 5.6%, while the pre-op has an error of 1.3%. There is no explanation as to how Yasuda et al. [208] handle the foramen magnum; apparently it is closed by letting the user manually remove the lower slices and thus putting a constraint on the skull base.

With its combination of full automation, high speed and high accuracy, the new method developed and applied in Chapter 9 is believed to be superior to methods previously published. Due to the same advantages, it should also have a potential to be developed into a user-friendly tool that could be used in a clinical setting, either by a radiologist or a surgeon.

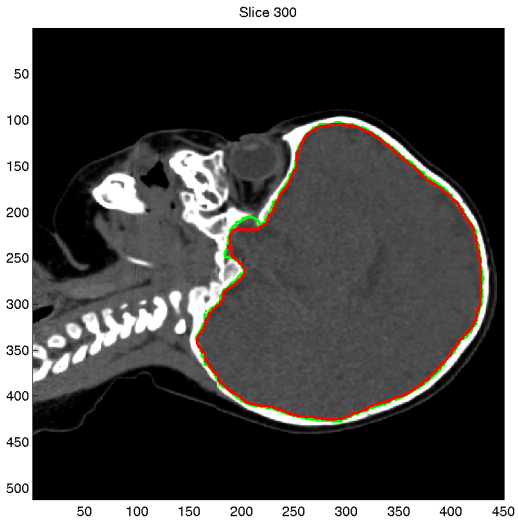


Figure 7.1: Manual contour in red and the automatic method in green. The automatic method seems to perform better in the skull base compared to the manual.

Future work is suggested to include a exhaustive investigation of the accuracy of the manual segmentation, in order to develop a gold standard to which other methods could be compared. In particular, such a gold standard should build upon a thorough decision of how to close foramina during manual segmentation. Figure 7.1 provides an example of a clearly non-optimal manual segmentation. Furthermore, it would be interesting to try the active surface approach proposed by Eskildsen et al. [61], where polygonal meshes are deformed to fit the boundaries

of the cerebral cortex in MR images using a force balancing scheme. A future goal would also be to determine the ICV's of the normal population as a function of age. This would be of high clinical importance as a reference standard for the ICV in patients.

In Chapters 10 to 12, non-rigid image registration has been applied to the study of craniofacial anomalies. This has involved the concept of a computational atlas as well as the analysis of deformation fields to estimate anatomical variability, determination of point correspondences, and growth modeling.

The analysis of deformation fields is highly dependent on the accuracy of the image registrations. Accordingly, validation of the registration is imperative. In a paper by Klein et. al from 2009 [97], 14 non-linear registration algorithms were evaluated and IRTK [162], which is the one used in the present thesis, was ranked as one of the best non-rigid registration algorithms. Validation relative to manual annotation carried out by experienced observers is usually recommended. It requires, in turn, that the intra- and inter-observer error is determined. Even though landmarks are recommended for validation [158], manually annotated landmarks are always limited by their sparse distribution compared to the voxel-wise correspondences achieved from non-rigid image registration. A common means of augmenting the point-to-point validation (as carried out e.g. in Chapter 10), is to determine point-to-surface errors (distance between every point in the surface and the closest location on the target), which gives a high spatial resolution of the registration errors and a lower bound for the point-to-point error.

In Chapter 10, non-rigid registration was applied to CT scans of a group of Taiwanese infants with UCLP in order to quantify the primary morphology (the morphology before any treatment). The results included the computation of the mean and variability of head shape at full spatial resolution, i.e. everywhere in the head and both for bone and soft tissue. It surpasses the spatial detail obtained in any previous studies, and demonstrates a very promising method for the future when low radiation dose volumetric CT-like scanners would be expected to become available. The developed method would open up for an automatic measurement of virtually any clinically relevant parameter (e.g. distance, angle, volume). Furthermore, the method developed in Chapter 10 was validated relative to manually placed landmarks, and seen to perform on equal terms with the manual method.

Different recent approaches for quantification of variability in groups of children with CLP are e.g. [32, 80]. Hermann [80]] used landmarks and cephalograms for quantification of morphology and growth in infants with CLP. Bugaighis et al. [32] acquired 3D facial scans of various types of cleft lip and/or palate patients after surgery and applied Procrustes Analysis and Principal Components

Analysis (PCA) to investigate differences in size and shape within and between the groups. In addition, TPS was utilized to allow interactive visualization of the differences. The last-mentioned approach is interesting to apply to our dataset as it has similarities to 3D facial scans when surfaces are extracted. Preliminary results using PCA showed that 12-13 modes were needed in order to retain 95% of the total variance. While Parallel Analysis, a tool for determining the number of retainable modes from e.g. a PCA, showed that only 5-6 modes were not considered as random noise.

In Chapter 11, non-rigid registration was tested in order to determine detailed point correspondence between CT scans of the same individual with UCLP before and after lip adhesion. A successful determination of the deformation vectors between the two stages would allow quantification of the change due to surgery, as well as providing a basis for surgery evaluation. As expected, the differences in morphology before and after lip adhesion turned out to be too large for a successful use of non-rigid registration alone. However, in Chapter 11, prior information about the typical result of lip adhesion was included in order to guide the non-rigid registration, and this was seen to improve the result.

Chapter 12 examines another difficult problem: the non-rigid registration between mouse skulls of different ages. Hildur Ólafsdóttir [105] demonstrated the successful application of atlas based non-rigid registration to study the differences in morphology of a group of adult Crouzon mice and a group of normal mice. In Chapter 12 the challenge was to register the adult mice (6 weeks old) to younger (4 week old) mice in order to study the differences in growth patterns between Crouzons and normals. Previous studies on growth are, e.g. mandibular growth using landmarks [9, 82]. As with morphology, the atlas method was able to provide deformation vectors at every voxel, in this case allowing convincing side-by-side animations of Crouzon and normal growth, respectively, demonstrating many similarities with growth patterns in humans. Again it was demonstrated that this methodology represented a convincing tool for automated phenotyping: a tool that would be expected to become increasingly important as modern animal scanners deliver increasingly high quality images of the many animal models being developed.

A limitation of the study of mouse growth was the small number of mice included (5 mice in each of the 4-week-old groups; 10 mice in each of the 6-week-old groups). However, the data set was sufficiently large in order to demonstrate the feasibility of the approach, including the methods for determining statistical significance (local false discovery rate and MANOVA). A future goal would be to increase the sample size in order to increase the effective signal to noise ratio of the result.

Another possibility for increasing the signal to noise ratio would be to increase the time span between groups, thus increasing the signal (longer deformation vectors as more growth has taken place). However, there would be a risk that differences would become too large for the non-rigid registration to be able to successfully determine point correspondences. A better solution would be to include more time steps (increase temporal resolution). In fact, the mouse data studied in Chapter 12 contains more micro-CT scans at more ages than those selected for study in the present thesis: there are groups of mice scanned 2 days before birth, then at 1 week, 2 weeks and 3 weeks of age in addition to the 4 and 6 weeks used here. Studying the growth across all these ages presents, however, a new problem concerning the ability of the method to deal with topological changes. During the early stages of growth, bones of the skull merge, thus changing the topology of the dataset; a situation not well handled by non-rigid registration.

An alternative way to model growth could be to model each bone part (as an average) with the proposed method. In the case of a particular bone missing due to the developing state, this bone could be left out. Also, a model constructed from longitudinal data would be interesting as each mouse/subject could be modeled. A second possible way to deal with the topological changes (in the Crouzon mouse datasets, but also in the UCLP datasets of Chapter 11) is to extract iso-surfaces from the volumetric data and turn to computer graphics. Misztal [125, 126, 127] proposed a method, Deformable Simplical Complexes (DSC), for dealing with topological changes in his PhD thesis. The DSC method relies on a 3D interface, i.e. a surface which separates different types of material, for instance bone and soft tissue. The DSC method has a number of advantages compared to the level set method [135, 184] or adaptive snakes [123]: Firstly, points are easy to track as the surface only changes as necessary when pieces collide. Secondly, tetrahedra/polygonized surfaces are used, and these are easily obtained for medical data e.g. using Marching cubes [114]. Finally, the method might be extendable to two or more phases e.g. several types of tissues.

7.1 Conclusion

This thesis has demonstrated the feasibility, accuracy and efficiency of applying non-rigid image registration and image segmentation using graph cuts to study craniofacial anomalies. The methods and results are presented through publications in both the technical and medical domains.

As shown in Table 7.1, the objectives and requirements of the PhD-study have been met. Each of the four studies (Chapters 9-12) has provided new knowledge

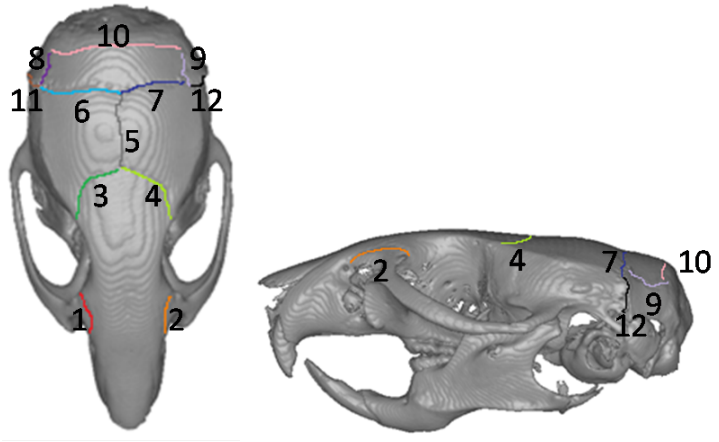
in the area of craniofacial research, in addition to providing image analysis tools that have the potential of being used not only in the image analysis laboratory for population studies, but also in the clinic for estimation of relevant parameters in order to provide improved diagnostics, treatment planning and treatment evaluation to the patients.

Chapter 8

Appendix

8.1 Suture Ratings

The sutures of the 4 and 6-weeks Crouzon mice were rated by an expert on a five-point scale from closed to full open. The 12 sutures are illustrated on a 6-week normal mouse in Figure 8.1, and the ratings are given in Table 8.1. The IDs 1-5 correspond to the IDs on the 4-weeks, while IDs 11-20 correspond to the 6-weeks.



(a) Top view. (b) Lateral view.

Figure 8.1: Wild-type mouse labeled with the 12 sutures. Suture 1 (red), 2 (orange), 3 (dark green), 4 (light green), 5 (gray), 6 (light blue), 7 (dark blue), 8 (dark purple), 9 (light purple), 10 (rose-pink), 11 (brown) and 12 (black).

Suture number	Closed	Almost closed	Partially open	Almost open	Open
1	2,all 6-weeks				1,3,4,5
2	4,5,all 6-weeks		1		2,3
3	1,3,5,11,12,13,14, 15,16,17,18,19	2			4,20
4	2,3,5,12,13,14,15, 16,17,18,19	1,20		4,11	
5	14,16,17,18,19	14,15	5	1,13	2,3,4,11,20
6	15,17,18,19		12	3,13,16,20	1,2,4,5,11,14
7	15,17,18,19		11,16,20	3,12	1,2,4,5,13,14
8	1,2,3,5, all 6-weeks		4		
9	5,all 6-weeks		1,2,3,4		
10	All 4-weeks, all 6-weeks				
11	3,all 6-weeks				1,2,4,5
12	11,12,14,15,16, 17,18,19,20				All 4 weeks, 13

Table 8.1: Suture ratings for Crouzon mice: Closed, almost closed, partially open, almost open and fully open refers to the suture condition. The numbers 1-5 denote 4-week, while 11-20 denote 6-week.

8.2 UCS data

ID	Diag.	Age [Days]	Resolution [mm] x	y	z	Slices in z	Intensity range [Min; Max]
1	LUCS	129	0.37	0.37	1.00	178	[-463; 3030]
2	RUCS	127	0.36	0.36	1.00	167	[-1351; 2021]
3	RUCS	130	0.34	0.34	0.80	247	[-3488; 2418]
4	RUCS	162	0.39	0.39	0.90	189	[-3428; 2378]
5	LUCS	413	0.43	0.43	1.00	184	[-1074; 2892]
6	RUCS	444	0.39	0.39	0.60	350	[-7664; 22432]
7	LUCS	206	0.37	0.37	1.00	311	[-2275; 3470]
8	LUCS	157	0.49	0.49	0.80	226	[-3446; 2721]
9	LUCS	289	0.39	0.39	0.80	251	[-3446; 2711]
10	RUCS	136	0.39	0.39	0.80	241	[-3442; 3506]
11	LUCS	483	0.47	0.47	0.80	226	[-3486; 3133]
12	RUCS	306	0.43	0.43	0.65	289	[-3428; 2814]
13	LUCS	159	0.39	0.39	1.00	178	[-3437; 2487]
⋆14	RUCS	437	0.45	0.45	0.60	313	[-1100; 2917]
15	LUCS	257	0.43	0.43	0.60	325	[-3441; 2824]

Table 8.2: Facts about the UCS data. Patient #14 was acquired in Helsinki. The diagnosis LUCS and RUCS denotes left and right UCS, respectively.

8.3 UCLP data

ID	Diagnose	Pre-op intensities [Min; Max]	Post-op intensities [Min; Max]	Age at 1st CT (Pre-op) Days [months]	Age at 2nd CT (Post-op) Days [months]
1	L-UCLP	[0; 168]	[0; 217]	90 [3,0]	305 [10,2]
2	R-UCLP	[0; 179]	[0; 215]	87 [2,9]	341 [11,4]
⋆3	?	[0; 178]	[0; 209]	?	?
4	R-UCLP	[0; 213]	[0; 184]	94 [3,1]	295 [9,8]
5	R-UCLP	[0; 171]	[0; 192]	85 [2,8]	319 [10,6]
⋆6	L-UCLP	[0; 255]	[0; 254]	90 [3,0]	311 [10,4]
7	L-UCLP	[0; 255]	[0; 208]	102 [3,4]	332 [11,1]
8	L-UCLP	[0; 255]	[0; 255]	83 [2,8]	324 [10,8]
9	L-UCLP	[0; 255]	[0; 254]	80 [2,7]	321 [10,7]
⋆10	R-UCLP	[0; 254]	[0; 192]	93 [3,1]	313 [10,4]
11	L-UCLP	[0; 254]	[0; 255]	89 [3,0]	342 [11,4]
12	L-UCLP	[0; 255]	[0; 254]	88 [2,9]	316 [10,5]
13	R-UCLP	[0; 254]	[0; 253]	87 [2,9]	321 [10,7]
⋆14	L-UCLP	[0; 255]	[0; 167]	92 [3,1]	335 [11,2]
15	L-UCLP	[0; 253]	[0; 189]	90 [3,0]	353 [11,8]
16	R-UCLP	[0; 255]	[0; 199]	89 [3,0]	295 [9,8]
17	R-UCLP	[0; 253]	[0; 189]	93 [3,1]	356 [11,9]
18	L-UCLP	[0; 254]	[0; 214]	87 [2,9]	331 [11,0]
⋆19	R-UCLP	[0; 250]	[0; 197]	102 [3,4]	331 [11,0]
20	L-UCLP	[0; 254]	[0; 215]	20 [0,7]	343 [11,4]
21	L-UCLP	[0; 255]	[0; 186]	96 [3,2]	335 [11,2]
22	L-UCLP	[0; 173]	[0; 201]	37 [1,2]	293 [9,8]
23	L-UCLP	[0; 175]	[0; 207]	117 [3,9]	283 [9,4]

Table 8.3: Overview of the UCLP data. L-UCLP and R-UCLP, denote left and right UCLP, respectively. No information about patient #3 was available. ⋆ denotes the excluded patients. Patient #1 was later excluded to avoid potential bias as this patient was used as target for atlas construction.

Part II

Contributions

Chapter 9

Genus Zero Graph Segmentation: Estimation of Intracranial Volume

Rasmus R. Jensen[†], Signe S. Thorup[†], Rasmus R. Paulsen[†], Tron A. Darvann[‡,⋆], Nuno V. Hermann[‡,♯], Per Larsen[‡], Sven Kreiborg[‡,♯,◇], Rasmus Larsen[†]

[†)] DTU Informatics, Technical University of Denmark, Lyngby, Denmark
[‡)] 3D Craniofacial Image Research Laboratory, Copenhagen, Denmark
[⋆)] Dept. of Oral and Maxillofacial Surgery, Copenhagen University Hospital, Copenhagen, Denmark
[♯)] Pediatric Dentistry and Clinical Genetics, University of Copenhagen, Denmark
[◇)] Dept. of Clinical Genetics, Copenhagen University Hospital, Copenhagen, Denmark

ABSTRACT

The intracranial volume (ICV) in children with premature fusion of one or more sutures in the calvaria is of interest due to the risk of increased intracranial pressure. Challenges for automatic estimation of ICV include holes in the skull e.g. the foramen magnum and fontanelles. In this paper, we present a fully automatic 3D graph-based method for segmentation of the ICV in non-contrast CT scans. We reformulate the ICV segmentation problem as an optimal genus 0 segmentation problem in a volumetric graph. The graph is the result of a volumetric spherical subsample from the data connected using Delaunay tetrahedralisation. A Markov Random Field is constructed on the graph with probabilities learned from an Expectation Maximisation algorithm matching a Mixture of Gaussians to the data. Results are compared to manual segmentations performed by an expert. We have achieved very high Dice scores ranging from 98.14% to 99.00%, while volume deviation from the manual segmentation ranges from 0.7%-3.7%. The Hausdorff distance, which shows the maximum error from automatic to manual segmentation ranges, from 4.73-9.81mm. Since this is sensitive to single error, we have also found the 95% Hausdorff distance, which ranges from 1.10-3.65mm. The proposed method is expected to perform well for other volumetric segmentations.

Index Terms— Intracranial volume, CT, craniosynostosis, graph cut, segmentation

9.1 Introduction

Unicoronal synostosis (UCS) is a congenital craniofacial malformation characterized by the premature fusion of one of the coronal sutures, potentially leading to asymmetric head shape, craniofacial growth disturbances, increased intracranial pressure and developmental delays. Computed Tomography (CT) scanning is usually performed to confirm the diagnosis and to facilitate surgical treatment planning. The intracranial volume (ICV) in children with premature fusion of one or more sutures in the calvaria may become reduced, leading to risk of increased intracranial pressure [205]. Challenges for automatic estimation of ICV include holes in the skull in newborns (the fontanelles), but also holes in the cranial base (e.g. the foramen magnum and other foramina, fissures and synchondroses). The main contribution of our work is a fast and fully automatic method for segmentation and estimation of the ICV in CT scans of children with craniosynostosis. The method is based on the construction of a volumetric graph description of the skull volume using tetrahedralization followed by a graph cut

forced to robustly perform a genus 0 segmentation. Validation is carried out by comparing the automatic segmentation model to a semi-automated model.

9.2 Brief Review of the Previous Research

Current work on automatic ICV[1] estimation has focused on Magnetic Resonance Imaging (MRI) volumes [120, 139, 180]. However, these methods are not well suited for ICV estimation in craniosynostotic cases due to the limited bone-tissue contrast in MRI. In the case of craniosynostosis, the best contrast of the cranial bones, e.g. for diagnosis and surgery planning, is obtained from CT scans. Furthermore, standard methods often use atlases based on a normal population, which may lead to a bias in the estimation of the ICV in craniosynostotic cases. The current standard for ICV estimation from CT is a manual method based on thresholding followed by a seed-growing algorithm. The challenge of this method is the need for manual editing in the various foramina in the skull base as well as in regions where craniosynostosis or lacking suture fusion have caused gaps between the cranial bones [176, 7, 8].

Anatomical segmentation such as the segmentation of the ICV in medical images is addressed in the literature by a series of approaches. In [197], deformable template matching is applied in a Bayesian setting; in [112], deformable surface models are proposed using a graph cut approach; and in [211], a multiclass Markov Random Field (MRF) is used for voxel classification. In the latter case it is interesting that, for two-class models, global optimal segmentation can be obtained using a graph-cut-based approach [101]. In this work we propose a two class segmentation of the ICV, where the classes (inside and outside) are modeled as mixtures of Gaussians. In addition to a label prior, we use a gradient-dependent interaction term. Moreover, we employ a tetrahedralization of a spherical equidistant sample distribution leading to a graph. The graph has dedicated outside and inside nodes, which robustly forces the graph segmentation to be of genus 0.

9.3 Approach

The data consist of pre-surgical CT head scans of 15 children diagnosed with UCS (either left- or right-sided). Age ranged from 6 to 18 months. All scans were acquired at Copenhagen University Hospital, Rigshospitalet, except for one

[1] In MRI they often estimate the total intracranial volume (TIV).

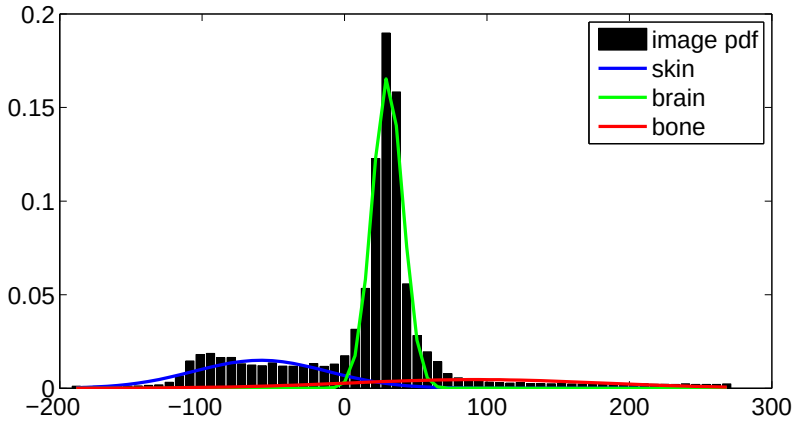


Figure 9.1: A mixture of Gaussians shown on the intensity histogram of a CT scan.

which was acquired at Helsinki University Central Hospital. Because of the UCS and the different age the data set is not homogeneous. All scans were obtained at 512 x 512 pixels in-plane size and a complete volume consists of between 167 and 350 slices.

The aim of the method is to create a volumetric segmentation that follows the transition between brain matter and bone, while also closing holes in the bone structure. As the intensities of the CT scans vary, we fit a mixture of Gaussians to each individual scan. The mixture of Gaussians is carried out using expectation maximization and results in three normal distributions describing: skin, brain matter and bone (see Fig. 9.1). Skin and bone have higher variance compared to brain tissue, which is used to classify the distributions unsupervised. Brain matter is by far the dominant, but also that with the least variance. Using the probability density function, where v is a sample value, we define the following two probabilities: $\mathrm{p}(v|x = \mathrm{ICV}) = \mathrm{pdf}_{brain}$ and $\mathrm{p}(v|x \neq \mathrm{ICV}) = \mathrm{pdf}_{skin} + \mathrm{pdf}_{bone}$. Generally, the brain matter distribution fits well to the data, while the other two tissues just stay below and above both with a wider standard deviation. Using only the mixture of Gaussians to classify brain-tissue and non-brain tissue would lead to misclassification as the distributions are crude, while the proposed method is insensitive to this.

Before the segmentation, the volumes were interpolated in the slice-wise direction to create isotropic voxels and ensure a regular sampling. A graph is created on sampling points in the volumes. The sampling points are found using a spherical volume of quasi-equidistantly distributed nodes. The nodes are distributed on

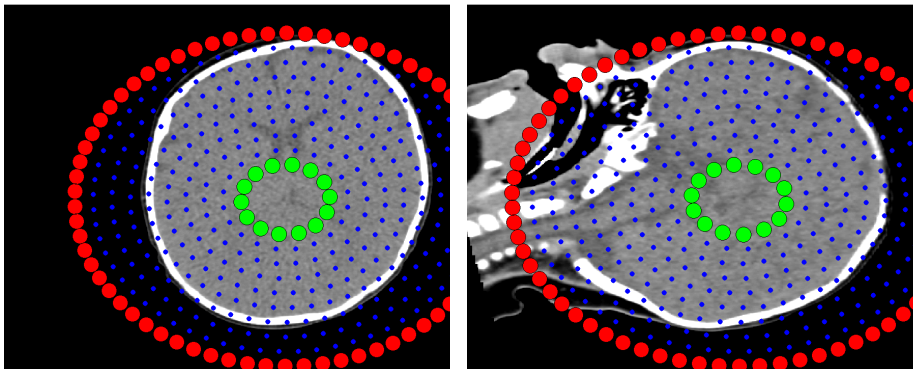


Figure 9.2: Examples of distribution of sample points in the transverse and sagittal planes. The sample density is much higher in the actual application with a voxel distance of two between sample points. The green nodes show the inner sphere which is forced to be part of the ICV, while the red nodes shows the outer sphere which is forced to be outside the ICV. The slices are contrast enhanced based on the mixtures of Gaussians.

the surfaces of concentric spheres, where the differences between their radii are equal to the spacing between their longitudes. Similarly, points on the longitudes are distributed with this spacing. This suggested sampling approach has two benefits: the sampling density can easily be changed for different resolutions, and it also removes the over/under-sampling problem of a spoke-like graph directed from the center and out. The sample volume is centered in the middle of the calvaria and created such that it covers the entire skull. In the center of the graph, we leave a small empty sphere, which will be used to clamp the inside of the graph to the ICV (see Fig. 9.2). The spherical graph is centered automatically by summing the voxel-wise probability of brain matter given the distribution prior. We find the coordinates of voxels with higher probability than $\frac{2}{3}$ of the maximum summation of the sagittal, coronal and transverse planes, respectively. The center is found as the median of these coordinates, while the radius of the sphere is found as 2.5 times the maximum interquartile range in the sagittal and coronal planes only. Fig. 9.2 illustrates the sample point distribution in the transverse and sagittal planes. For the actual ICV estimation we have used a much higher sample density, using an even voxel distance of two between sample points. A robust way of connecting each of the sample points to the immediate spatial neighborhood in a highly connected graph can be achieved by Delaunay tetrahedralization [178]. As this approach produces doublets of edges represented by several adjacent tetrahedra the connectivity has to be cleaned up such that edges are represented only once. On the graph with index set $\mathcal{I}$, we

define the following MRF, which is solved using graph cuts [101]:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{I}} \left(\Phi(v_i|x_i) + \sum_{j \in N_i} (\lambda(x_i, x_j) + \psi(x_i, x_j)) \right) \tag{9.1}$$

$\Phi(v_i|x_i) = -\log \mathrm{p}(v_i|x_i)$ defines a log-likelihood function. The function returns high values for low probabilities and vice versa. The outer and inner sphere log-likelihood values are clamped as follows:

$$\begin{aligned} \Phi(i \in \text{outer}|x_i) &= \begin{cases} \infty & x_i \neq \text{ICV} \\ 0 & x_i = \text{ICV} \end{cases} \\ \Phi(i \in \text{inner}|x_i) &= \begin{cases} 0 & x_i \neq \text{ICV} \\ \infty & x_i = \text{ICV} \end{cases} \end{aligned} \tag{9.2}$$

N_i denotes the neighborhood of the i'th voxel and has terms defined as:

$$\lambda(x_i, x_j) + \psi(x_i, x_j) = \begin{cases} K_\nabla e^{-|\nabla \mathrm{f}(x_i, x_j)|} + K_{ij} & x_i \neq x_j \\ 0 & x_i = x_j \end{cases} \tag{9.3}$$

Where $|\nabla \mathrm{f}(x_i, x_j)|$ is the absolute gradient; K_∇ controls the power of the gradient term, while K_{ij} is a general smoothness prior. With the outside sphere clamped as *outside ICV* and the inside sphere clamped as *ICV*, a setting of K_∇ and K_{ij} exists for which the resulting graph cut will only be on the inner edge of the skull, close the holes (e.g. fontanelles, optic canals, foramen magnum, and other foramina), and it will produce a segmentation of genus 0. We achieved our results with K_∇ being of the same magnitude as $\max \Phi(v_i|x_i = \text{ICV}), i \in I$ and K_{ij} $\frac{1}{100}$ of that. We found the results to be rather insensitive to fine tuning of K_∇ and K_{ij}. The result of a segmentation is shown in Fig. 9.3. With a segmentation of the graph, the volume can be estimated using the tetrahedralization of the sample points contained in the cut.

9.4 Results and Discussion

Ground truth was made by expert manual segmentations using a semi-automatic, slice-wise method based on a seed-growing algorithm incorporated in Analyze™ (BIR Research Lab, Mayo Clinic, Rochester, MN, USA). This method requires a user-specified intensity threshold and manual editing.

As gaps and small fractures are present in the data, the semi-automatic segmentation algorithm often breaks down and manual editing is necessary. Easy cases for manual editing are when the natural curvature of the skull is present and the gaps are small. Unfortunately, severe cases with large gaps and no a priori information

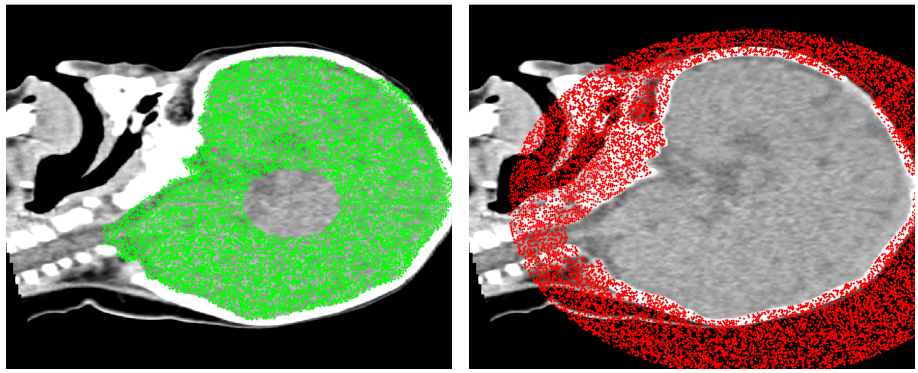


Figure 9.3: The result of the segmentation between *ICV* and *outside ICV* shown in the sagittal plane. The slices are contrast enhanced based on the mixtures of Gaussians.

potentially lead to large errors. Average processing time for the manual method is two hours, including threshold estimation. Average runtime including all steps of the process is 12 minutes on a fast consumer desktop computer (Intel i7 3.6@4.2 GHz processor with 16 GB ram) running Matlab. Figure 9.4 illustrates the two methods on example slices. Looking at the transversal slice, the methods are consistent, while on the sagittal slice the automatic method includes part of the optic canal. Fig. 9.5 shows the automatically estimated volumes as a function of the manual volumes. The linear regression of the two lines has been forced through origo. For the volume measured entirely inside the cut, we get R^2-value of 99.54%, with a slope of $\alpha = 0.9768$ being a 2.32% underestimate. Including the tetrahedra partially inside the cut, we get and R^2-value of 99.50%, with a slope of $\alpha = 1.0246$ being a 2.46% overestimate. We have used the *worst* of the volume estimates (i.e. including the partially cut tetrahedra) to calculate volume deviation, Dice score [52] and Hausdorff distance [85].

Table 9.1 shows the comparison between the two segmentation models. For each patient the deviation from the manual volume, Dice coefficient and Hausdorff distance were calculated to evaluate the proposed method. While the Dice coefficient measures the volume overlap, the Hausdorff distance measures the maximum error between the two segmentations. Since the Hausdorff distance measure is sensitive to single error the 95 % Hausdorff distance is also included.

Deviations in volume are very small and lie between 0% and 3.7%. The Dice coefficients also show a high consistency in overlap between the methods. The lowest Dice score is 98.14, while the highest is 99.00. An explanation for the differences might relate to the graph cut lying slightly outside the manual border

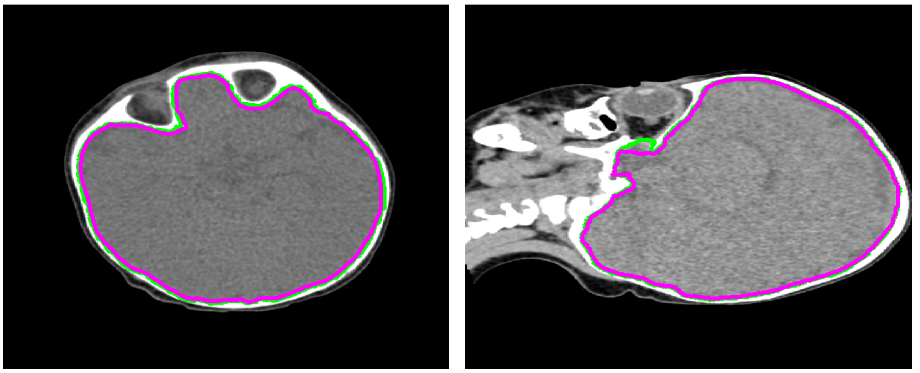


Figure 9.4: Examples of the manual (magenta) and automatic segmentation (green) shown on a transversal (Patient #8) and sagittal slice (Patient #11), respectively. The slices are contrast enhanced based on the mixtures of Gaussians.

(see Fig. 9.4). This behavior might relate to the chosen threshold, and it would be interesting to assess the consequence of using various thresholds as well as the manual error in future work.

The Hausdorff distance, which is a conservative measure, showing the maximum error in overlaps, shows differences of up to 9.81 mm. Examining the images, large differences between the two methods occur where the foramen magnum and optic canals are closed. Both regions are hard to segment consistently, since the spinal cord and the optic nerves are similar in intensity to that of brain matter.

9.5 Concluding Remarks

In conclusion, we have implemented an automatic, fast method for accurate estimation of the ICV in children with UCS. The method is fairly insensitive to fine tunning of parameters. There is good reason to believe that the proposed method can be used for other applications in volumetric image segmentation.

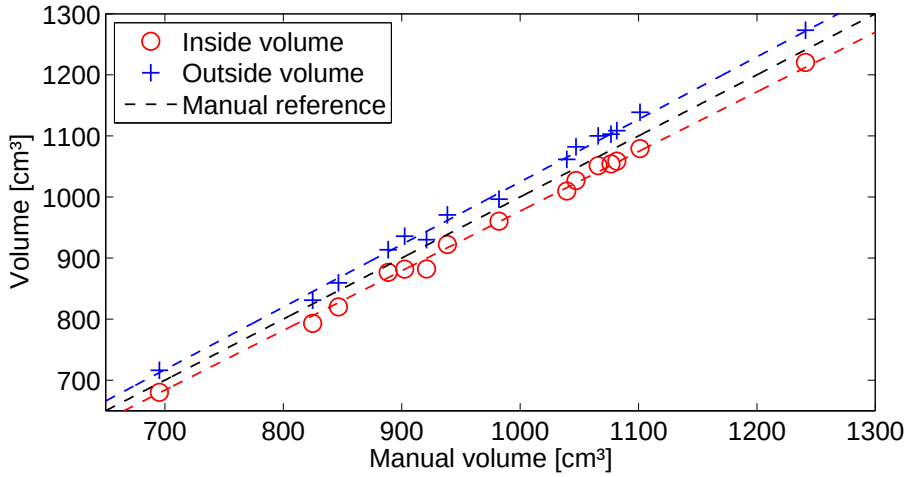


Figure 9.5: Automatic volumes as a function of the corresponding manual volume including linear regression lines. Red denotes the volume with tetrahedra fully inside the segmentation, while blue also includes tetrahedra partially inside. As a reference the manual volume is shown as a black line.

Patient #	Manual volume $[cm^3]$	Automatic volume $[cm^3]$	Volume deviation [%]	Dice score [%]	Hausdorff distance $[mm]$	95% Haus. distance $[mm]$
1	846.6	859.5	1.5	98.87	7.62	2.22
2	824.8	831.0	0.7	98.88	5.09	2.16
3	888.6	913.5	2.8	98.55	4.94	1.16
4	1076.9	1102.9	2.4	98.67	5.30	1.10
5	1039.8	1061.7	2.1	98.71	6.88	1.72
6	1241.1	1273.2	2.6	98.66	4.74	1.10
7	982.2	996.5	1.5	99.00	9.81	3.65
8	902.6	935.8	3.7	98.14	4.88	1.38
9	1081.4	1108.8	2.5	98.67	4.73	1.12
10	1066.1	1100.0	3.2	98.36	7.97	1.75
11	1101.2	1138.8	3.4	98.28	6.05	1.34
12	921.1	930.2	1.0	98.87	8.11	2.98
13	695.2	716.2	3.0	98.39	6.09	1.89
14	1047.5	1082.1	3.3	98.28	6.48	1.27
15	938.6	970.8	3.4	98.29	5.16	1.22

Table 9.1: Result overview comparing the manual segmentation to the proposed automatic method. *Volume deviation* denotes the relative percentage-wise difference between the numeric volume estimates. *Dice score* measures the overlap between the volumes. *Hausdorff distance* assesses the maximum error. *95 % Hausdorff distance* is included as this measure is much less sensitive to single error.

Chapter 10

A Method for Automated 3D Quantification in CT Scans of Infants with Cleft Lip and Palate

Technical Report

Signe S. Thorup[†], Tron A. Darvann[‡,⋆], Nuno V. Hermann[‡,♯], H. Ólafsdóttir[†], Rasmus R. Paulsen[†], Per Larsen[‡], Alex A. Kane[□], L.-J. Lo[△], Rasmus Larsen[†], Sven Kreiborg[‡,♯,◇]

[†)] DTU Informatics, Technical University of Denmark, Lyngby, Denmark
[‡)] 3D Craniofacial Image Research Laboratory, Copenhagen, Denmark
[⋆)] Dept. of Oral and Maxillofacial Surgery, Copenhagen University Hospital, Copenhagen, Denmark
[♯)] Pediatric Dentistry and Clinical Genetics, University of Copenhagen, Denmark
[◇)] Dept. of Clinical Genetics, Copenhagen University Hospital, Copenhagen, Denmark
[□)] Dept. of Plastic and Reconstructive Surgery, Washington University School of Medicine,

St. Louis, MO, USA
[Δ)] Dept. of Plastic and Reconstructive Surgery, Chang Gung Memorial Hospital and Chang Gung University, Taipei, Taiwan

ABSTRACT

Cleft Lip and/or Palate is the most common congenital craniofacial malformation and is caused by lack of fusion of facial processes in the young fetus leading to clefting that may be either uni- or bilateral. The data consists of 3D computed tomography (CT) scans of 17 Taiwanese infants with unilateral cleft lip and palate. The infants were scanned before lip repair at the age of 3 months, and again after lip repair at the age of 12 months. Using non-rigid registration, image data from a patient is automatically transformed into the coordinate space of a computational atlas representing an average patient head shape. In this process, all knowledge built into the atlas is transferred to the subject, thus creating a personalized atlas. The knowledge built into the atlas is e.g. location of anatomical regions and landmarks of importance to surgery planning / evaluation or population studies. With these correspondences various analyses could be carried out. The mean primary morphology in UCLP was determined as the mean of all the 17 datasets. Anatomical variability was determined as the mean and standard deviation of the distances from all patients to the atlas. Moreover, landmark-based measures such as cleft height, cleft width, nose angle and nose area was determined automatically and compared to the manual approach. We validated the registration using landmarks and point-to-surface errors.

Index Terms— Craniofacial growth, image registration, computational atlas, landmarks, morphological changes, cleft lip and palate.

10.1 Introduction

Cleft lip and palate (CLP) is the most common congenital craniofacial malformation with an incidence of 1:500 live births [81]. It is caused by incomplete fusion of the maxillary and palatal processes early in the embryonic life [116]. CLP can be uni- (UCLP) or bilateral (BCLP) and may involve soft tissues (lip and/or soft tissues in the alveolus and/or palate) and/or bone (alveolus and/or hard palate).

CLP is typically treated by surgical closure of the lip and palate during the first two years of life. The first surgical procedure takes place at 3-5 months of age and aims at closing the cleft of the lip. Many surgical approaches exist, but the so-called lip adhesion [116] is a common way of closing the cleft by stretching the soft tissue of the lip across the cleft. In the case of an additional clefting of the bone (alveolus and/or hard palate), the forces from the lip tissue stretching will also lead to a slight narrowing of the bony cleft during the time from lip adhesion

to the time of surgical closure of the palate which takes place approximately a year later.

3D volumetric imaging like CT and MRI is essential in order to precisely describe deviant craniofacial morphology and guide treatment management [15]. Detailed 3D information about pre-surgical morphology is important for planning the closure of the lip in CLP, as well as for planning subsequent surgeries [173]. Furthermore, it facilitates documentation and outcome evaluation.

This article presents a method for automated analysis of 3D volumetric images of individuals with CLP. The method is demonstrated on CT data of infants with UCLP obtained prior to lip adhesion as part of a former treatment protocol at Chang Gung Memorial Hospital, Taipei, Taiwan. Currently, the follow-up treatment protocol for the present patients has shifted towards the use of Cone Beam CT (CBCT) as the patients are older now. CBCT opens up for possible regular use of 3D volumetric imaging in CLP [198, 207, 77] due to reduction of the X-ray radiation dose relative to conventional CT. CT has been in limited use for assessment of the alveolar cleft [37], assessment of bone resorption after secondary alveolar bone grafting [63] and postoperative evaluation of adults with CLP [111].

Non-rigid image registration is commonly used to determine the spatial transformation that would bring two images of comparable anatomy into detailed spatial correspondence. The correspondences provide a deformation field which can be used to study the differences between the two images; e.g. representing differences between two subjects or temporal change in the same subject. In the present study, non-rigid registration techniques are applied in order to create an atlas of the primary morphology in UCLP. The atlas represents an average UCLP head shape. Using non-rigid registration, image data from an individual UCLP infant is automatically transformed into the coordinate space of the atlas, and in the process, all knowledge built into the atlas is transferred to the individual, thus creating a personalized atlas. The knowledge built into the atlas is e.g. location of anatomical regions and landmarks of importance to surgery planning / evaluation or population studies. A successful transformation thus fully automates the process of landmarking and measurement of clinically relevant parameters in the image data. The atlas method has been applied in a variety of medical fields, e.g. in the study of brain morphology (e.g. [34]) breast cancer screening (e.g. [134]) or cardiology (e.g. [55]). The application of the atlas method to the craniofacial region has so far been limited; although Christensen et al. [38] pioneered a method based on elastic transformations and applied it to segmentation of cranial CT data of children with craniofacial malformations. Later works are notably those by Lam et al. [106] and Cevidanes et al. [36].

In the present work, an atlas was built using non-rigid registration modeled by B-spline deformation guided by a normalized mutual information image similarity measure [163, 187, 162, 170]. The method has previously been applied to micro-CT scans of mice [133, 193] and CT scans of infants with unilateral coronal synostosis [132].

The method, being fully automatic, has potentially a multitude of applications. Firstly, it is able to present detailed information about the mean and variation of shape in the pre-surgical UCLP population. The richness of data compared to previous X-ray cephalometric studies (e.g. those using the 3-projection cephalometric X-ray method developed by Hermann et al. [81]) opens up for investigations asking questions about the cause of UCLP and other types of cleft lip and/or palate. Is the malformation in cleft lip and palate spatially isolated to the cleft region, and what are the anatomical differences between various types of CLP? Furthermore, clinically relevant parameters (e.g. for surgery planning and outcome evaluation in individuals) may be extracted, and this present article demonstrates the automated quantification of clinically relevant anthropometric parameters describing the cleft region.

The non-rigid registration methodology is validated by comparison to manually placed landmarks.

10.2 Material

Data consists of CT scans of 17 Taiwanese infants with non-syndromic UCLP obtained in a clinical context at Chang Gung Memorial Hospital in Taipei, Taiwan. The use of CT was approved by a local ethics committee at the hospital. The infants were born between 1997 and 1999, and scanned before lip repair at the age of 3 months (mean: 87 days, min/max: [20; 117] days). In-slice spatial pixel resolution ranged from 0.35 to 0.49 mm, and slice distance varied from 1.5 to 3.0 mm. The scans were reformatted in Taiwan to $340 \times 340 \times 400$ with equal resolution of 0.5 mm.

CT scanning of infants calls for anesthesia to avoid motion artifacts and therefore the scanning takes place at a hospital. Therefore, CBCT is not possible to use on infants as most scanners demands an up-right position and that the children are able to sit still for approximately 15 seconds. Since an anesthetized infant needs to lie down a conventional CT scanner has been used.

In UCLP, the cleft can be either on the right side or on the left side of the face - the left sided cleft being the most common. Assuming that individuals with

left sided clefts and those with right sided clefts represent the same biological entity as a group (the same mechanisms are causing the cleft to either appear on the left or the right side), all data of individuals with right-sided clefts were mirrored as a pre-processing step in order to appear as having left-sided clefts. This step increased the effective sample size and was necessary in order for all scans to be sufficiently similar for the non-rigid registration to succeed.

10.3 Methods

The goal of image registration is to warp one image, the source, into the coordinate system of another image, the target, using a transformation T. A basic image registration algorithm requires the following: a transformation model, T, a measure of image similarity, a regularization term or similar to ensure a valid transformation, and an optimization method to optimize the similarity measure with respect to the transformation parameters. In this study, two different transformation types were used, an affine transformation, and a non-rigid deformation based on B-splines [163, 162, 170]. The first captures global differences between the images while the latter handles local, nonlinear differences. In both types, normalized mutual information (NMI) [187] was used as a similarity measure and gradient descent optimization [163] was applied. The local support from using multilevel B-splines was enough for regularization.

10.3.1 Affine Registration

Affine registration applies an affine transformation to map the source image into the target. Affine transformation is a global transformation, defined by

$$T_{\text{global}}(x, y, z) = \mathbf{A}\mathbf{x} + \mathbf{t} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{10.1}$$

where, in 3D, $\mathbf{x}$ is a vector of pixel/voxel coordinates (x, y, z), $\mathbf{t}$ is a vector (t_x, t_y, t_z) representing the three translation parameters, and $\mathbf{A}$ is a matrix describing the rest of the parameters i.e. rotation, (anisotropic) scaling, and shearing [183, 75]. Due to the relatively small number of parameters being optimized, the registration is fast; however, only global differences between the images are taken into account while local differences are ignored. The affine registration is therefore used as an initialization step before the non-rigid registration is carried out.

10.3.2 Non-rigid Registration Based on B-splines

To obtain more detailed registration focusing on local differences, nonlinear transformations are required. A widely used method for this purpose is the non-rigid registration algorithm using B-spline-based free-form deformations (FFDs) [163, 170]. In this case, a composition of a global and a local transformation,

$$T(x, y, z) = T_{\text{global}}(x, y, z) + T_{\text{local}}(x, y, z) \tag{10.2}$$

is applied, where the global model consists of the affine transformation described above. In 3D, the local transformation model, the FFD, is defined by an $p_x \times p_y \times p_z$ grid of control points with spacing (δ_x, δ_y, δ_z). The underlying image is then deformed by manipulating the mesh of control points $\phi_{i,j,k}$. The FFD model can be written as the tensor product of one-dimensional (1D) cubic B-splines:

$$T_{\text{local}}(x, y, z) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \phi_{i+l,\, j+m,\, k+n}, \tag{10.3}$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $k = \lfloor z/n_z \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$ and $w = z/n_z - \lfloor z/n_z \rfloor$. B_0 through B_3 represent the basis functions of the B-spline:

$$B_0(u) = (1-u)^3/6, \tag{10.4}$$
$$B_1(u) = (3u^3 - 6u^2 + 4)/6, \tag{10.5}$$
$$B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6, \tag{10.6}$$
$$B_3(u) = u^3/6. \tag{10.7}$$

The transformation creates a dense deformation vector field which can be assessed at any point in the image.

10.3.3 Similarity Measure: Normalized Mutual Information

In order to bring images into correspondence by image registration, the degree of similarity between the two images needs to be defined. The normalized mutual information (NMI) is based on entropy measures in the two images. The marginal entropy in an image relates to the information content, or more intuitively it measures the uncertainty of guessing a voxel intensity. In image M with voxel

intensities $m \in M$ the discrete marginal entropy is defined as

$$H(M) = -\sum_{m \in M} p(m) \log p(m), \tag{10.8}$$

where $p(m)$ is the marginal probability. The joint entropy is defined on the overlapping region between the two images M and N with voxel intensities $m \in M$ and $n \in N$,

$$H(M,N) = -\sum_{m \in M, n \in N} p(m,n) \log p(m,n), \tag{10.9}$$

where $p(m,n)$ is the joint probability. This corresponds to the information content of the combined scene or the probability of guessing a pair of voxel intensities. Mutual information describes the difference between the sum of the marginal entropies and the joint entropy, and by dividing by the joint entropy, NMI is defined as

$$NMI(M,N) = \frac{H(M) + H(N)}{H(M,N)}. \tag{10.10}$$

The strength of entropy measures, such as NMI, is their ability to cope with two different modalities (e.g. [187]), but they have been widely used with good results in intra-modality applications as well (e.g. [161, 177]).

10.3.4 Atlas Construction

A computational, deformable 3D atlas was constructed in an iterative manner using non-rigid registration (a composition of an affine registration and B-spline-based non-rigid registration as described above). Figuratively, an atlas is similar to a 3D mean; a computational atlas represents an average UCLP head shape. Using non-rigid registration, image data from an individual UCLP infant is automatically transformed into the coordinate space of the atlas, and in the process, all knowledge built into the atlas is transferred to the individual, thus creating a personalized atlas. The knowledge built into the atlas is e.g. location of anatomical regions and landmarks of importance to surgery planning / evaluation or population studies. The atlas was constructed from the set of pre-op CT data and built according to the procedure listed in Algorithm 1. In line 5, the root-mean-square (rms) error between the voxel intensities of the current atlas and the previous atlas is calculated and an appropriate threshold value is chosen to define the state where the atlas stops changing. Lines 6 and 7 from Algorithm 1 are intended to reduce the bias in shape towards the choice of reference subject as previously done with good results [74, 161]. In this study we left out the reference subject from the subsequent analysis.

10.3.5 Quantitative Validation of Registration Accuracy

Two independent observers identified and annotated a number of pre-defined anatomical landmarks in all the $n = 17$ individual CT data. The landmarks were placed on surface representations of soft tissue and skull, respectively, extracted by the Marching Cubes algorithm [Lorensen1987] according to Figure 10.1. The landmark placement was carried out using a 3D software tool ('landmarker' by [48]) based on the Visualization Toolkit [172]. Also, the atlas was landmarked in the same way as the individual scans.

The average of the two annotations was used as a gold-standard (GS). The landmarks on the atlas were subsequently propagated automatically to all subjects using the previously obtained transformation. Finally, landmark errors were estimated. These were defined as the point-to-point error, i.e. the Euclidean distance from an automatically obtained landmark to the corresponding GS-landmark. The landmark errors were scaled to provide a reasonable comparison as described in the statistical analysis by [133] and Appendix Section 10.8.2.

Furthermore, the root mean square (rms) error was calculated as well as the error of the method, $s(i)$. The error of the method (Dahlberg's formula) is defined as

$$s(i) = \frac{1}{\sqrt{2}}\text{rms} = \sqrt{\sum\left(\frac{d^2}{2N}\right)}, \tag{10.11}$$

where $d = \|\bar{x}_1 - \bar{x}_2\|$,$\bar{x}_1$ and $\bar{x}_2$ is the coordinate of a landmark placed by observer 1 and 2, respectively, and N is the number of paired measurements [46]. $s(i)$ has been used by many orthodontic investigators, it represents the mean square error and follows from the addition theorem for the variance-distribution [76].

In addition to the point-to-point errors determined at each landmark, a much more spatially dense point-to-surface (closest point) error was determined by comparing the shape of each individual's surface to the shape of the deformed atlas. The atlas had been deformed to the particular individual's surface using the transformation obtained by the non-rigid registration. The metric of the comparison was the Euclidean distance from each point on the surface scan to the closest location on the deformed atlas.

Note that while the points on the individual's surface scan are defined by the surface mesh points given by the Marching Cubes algorithm, the closest location on the deformed atlas is often a position within the polygons.

It may be noted that the point-to-surface error typically under-estimates the true error (as more accurately represented by the point-to-point error). This is

due to the simplistic use of closest point locations in the definition of point correspondence: they do not necessarily represent true anatomical correspondence.

10.4 Results

The scaled errors from the quantitative validation of the non-rigid registration are shown in Figure 10.2 using box and whisker plots.

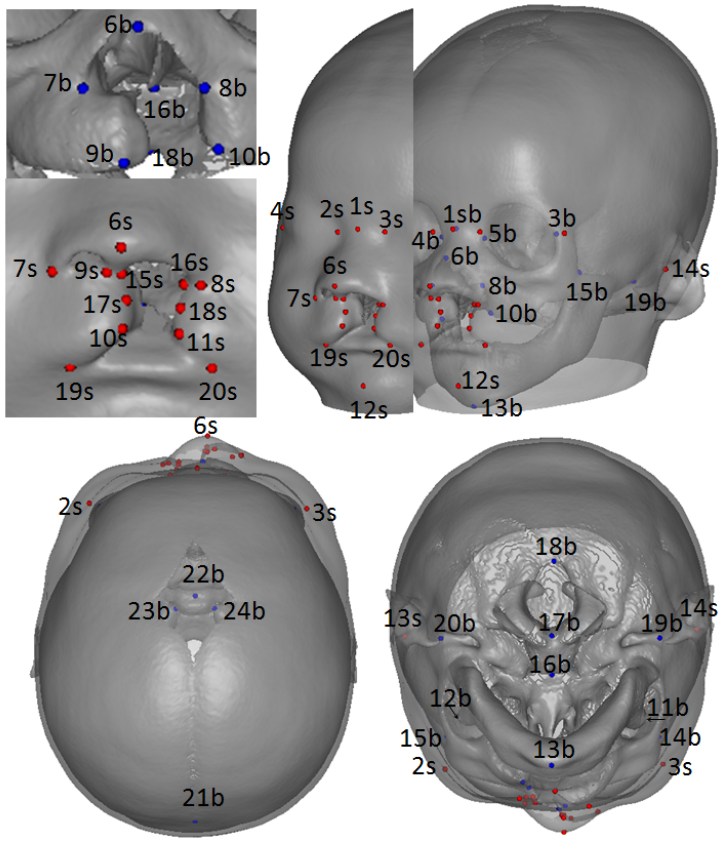
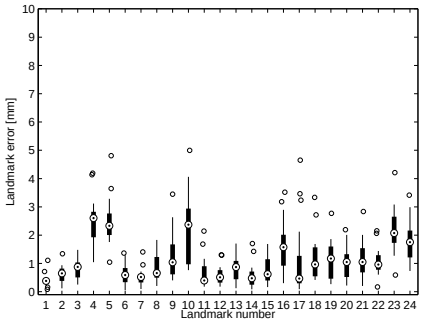


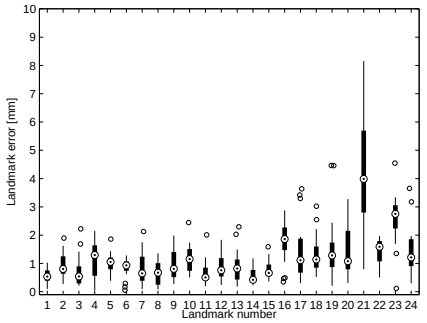
Figure 10.1: UCLP atlas with landmarks. Red landmarks labeled 1s-20s are soft tissue landmarks, while blue landmarks labeled 1b-24b are bone landmarks.

Table 10.1 shows the rms-error for soft tissue and bone, respectively.

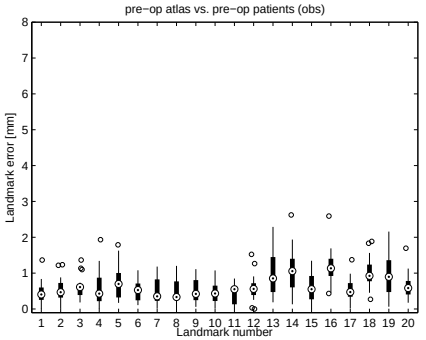
Table 10.2 shows the closest point differences, for both soft and bone tissue. The



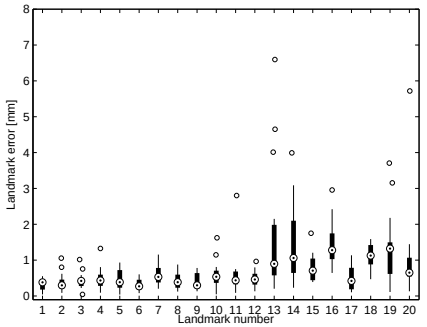
(a) Inter-observer error - bone



(b) Error of automatic method w.r.t. gold standard - bone



(c) Inter-observer error - soft tissue



(d) Error of automatic method w.r.t. gold standard - soft tissue

Figure 10.2: Landmark errors (in millimeters) based on the Euclidean distance. Left column: inter-observer errors (scaled by $\frac{1}{\sqrt{2}}$). Right column: landmark errors comparing gold-standard landmarks to the automatic method (scaled by $\sqrt{\frac{2}{3}}$). Top row: Bone landmarks. Bottom row: Soft tissue landmarks. The scaling factors are applied to obtain reasonable comparisons as explained in [133] and the Appendix Section 10.8.2. The following definition of a box- and whisker plot is used here: The box surrounds measurements between the upper and the lower quartile of the data. The symbol $\odot$ inside the box denotes the median of the data. The maximum length of the whiskers is 1.5 times the interquartile range (IQR). 'Outliers', those values lying outside the limits of the whiskers, are marked by a circle.

errors are the differences between the atlas warped to resemble each individual and the surface extracted from each individual CT-scan. The average method error (mean and standard deviation) for both kind of tissues are shown in Figure 10.3 and 10.4.

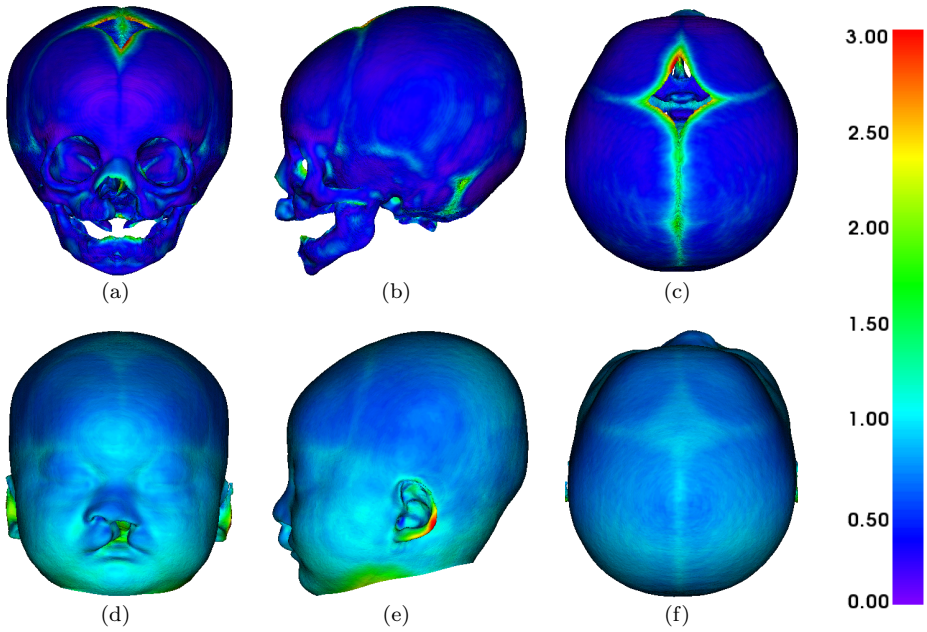


Figure 10.3: Average point-to-surface mean errors shown on the atlas in millimeters - top row bone tissue, bottom row soft tissue. Purple denotes no errors, while red denotes errors of 3 mm and above.

Figure 10.5 shows the variation in the dataset for both bone and soft tissue as the standard deviation from the mean.

Using the landmarks from Figure 10.1 the anatomical landmark-based measures: cleft width, cleft height, nose angle and nose area were estimated in Table 10.3.

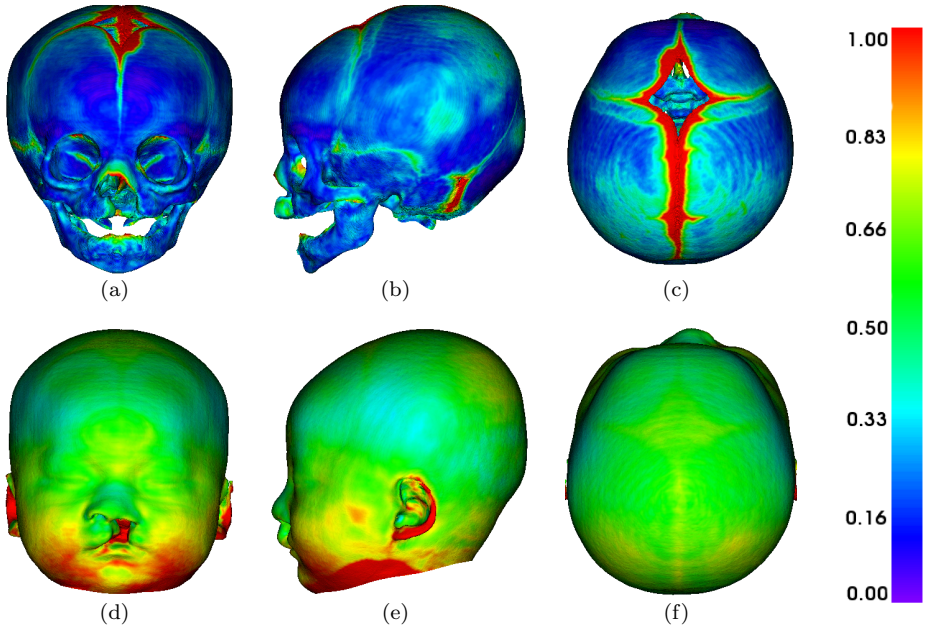


Figure 10.4: Average point-to-surface error standard deviations in millimeters w.r.t. the mean from Figure 10.3 shown on the atlas - top row bone tissue, bottom row soft tissue. Purple denotes no errors, while red denotes errors of 1 mm and above.

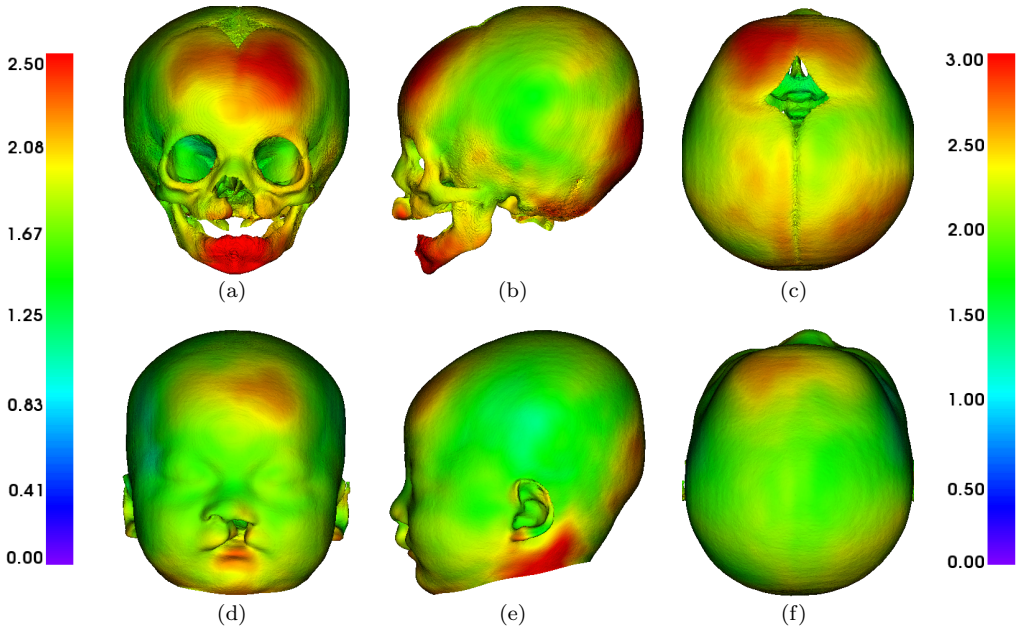


Figure 10.5: Bone and soft tissue atlases colored according to the standard deviation. The left colorbar belongs to the bone images, and the right colorbar belongs to the soft tissue images. Purple denotes no errors, while red denotes errors of >2.50mm/3mm, for bone and soft tissue, respectively. The colorbars are in millimeters.

Algorithm 1: Atlas construction

1: Atlas = a selected reference scan from the group
2: **repeat**
3: Register all scans from the given group to atlas
4: Atlas = Intensity average of all registered subjects
5: **until** atlas stops changing
6: Register atlas to all subjects from the given group
7: Transform atlas by T the average transformation obtained in step 6

Scaled	**Bone tissue**		**Soft tissue**	
rms [mm]	Manual	Automatic	Manual	Automatic
1	0,534	0,682	0,997	0,855
2	0,658	0,888	1,103	0,855
3	0,744	0,776	1,192	0,830
4	1,332	0,982	1,009	0,781
5	1,328	0,926	1,264	0,810
6	0,671	0,826	1,153	0,867
7	0,611	0,824	0,736	0,678
8	0,783	0,741	0,765	0,676
9	0,947	0,886	0,786	0,683
10	1,262	0,972	0,925	0,900
11	0,683	0,735	1,044	0,983
12	0,644	0,836	1,156	0,964
13	0,768	0,862	1,153	1,319
14	0,640	0,678	1,154	1,277
15	0,710	0,801	1,002	0,959
16	1,076	1,202	0,881	1,165
17	0,893	1,071	0,634	0,719
18	0,919	1,055	0,916	1,029
19	0,906	1,151	0,968	1,190
20	0,847	1,077	1,046	1,213
21	1,322	1,832	–	–
22	0,879	1,086	–	–
23	1,246	1,442	–	–
24	1,110	1,081	–	–
Average error	0,896	0,976	0,994	0,938

Table 10.1: Scaled rms-errors for bone and soft tissue landmarks. – indicates that the landmark is not defined.

Point-to-surface errors						
Individual	Bone tissue			Soft tissue		
1	0.522	±	0.438	0.193	±	0.207
2	0.454	±	0.399	0.549	±	0.239
3	0.442	±	0.340	0.246	±	0.272
4	0.399	±	0.342	1.733	±	0.504
5	0.617	±	0.458	2.127	±	1.146
6	0.467	±	0.344	0.861	±	0.401
7	0.462	±	0.309	0.307	±	0.256
8	0.603	±	0.441	1.160	±	0.275
9	0.477	±	0.383	0.907	±	0.325
10	0.335	±	0.310	0.395	±	0.195
11	0.425	±	0.300	1.189	±	0.595
12	0.502	±	0.425	0.237	±	0.229
13	0.426	±	0.362	0.829	±	0.380
14	0.483	±	0.754	0.787	±	0.393
15	0.516	±	0.341	0.662	±	0.342
16	0.474	±	0.361	0.385	±	0.580
17	0.455	±	0.537	0.829	±	0.342
Average	0.474	±	0.403	0.788	±	0.393

Table 10.2: Point-to-surface errors between the atlas warped to resemble each individual and the surface extracted from each individual CT-scan.

Symbol	Points	Mean ± std (manual)	Mean ± std (automatic)	p-value
$\leftrightarrow$ cleft width	{15s,16s}	13.194 ± 1.735	14.413 ±1.535	*0.038
$\leftrightarrow$ cleft width	{17s,18s}	11.667 ±1.386	12.527 ±1.476	0.09
$\leftrightarrow$ cleft width	{10s,11s}	13.576 ±1.521	13.212 ±1.268	0.454
$\leftrightarrow$ cleft width	{9b,10b}	19.409 ±2.235	19.091 ±1.932	0.66
$\leftrightarrow$ cleft height	{11s,16s}	8.156 ±1.339	8.583 ±1.484	0.386
$\leftrightarrow$ cleft height	{8b,10b}	9.215 ±1.084	8.855 ±1.218	0.371
$\angle$ nose	{1s, 6s, 9s}	58.433 ±5.651	58.708 ±5.269	0.884
Δ nose	{6s,7s, 8s, 6s}	139.646 ± 19.342	135.480 ± 21.801	0.56
Δ nose	{6b,7b, 8b, 6b}	138.425 ± 12.351	131.296 ± 12.271	0.101

Table 10.3: Landmark measures. $\leftrightarrow$, $\angle$ and Δ denotes distance, angle and area, respectively. Bone and soft tissue measures are abbreviated b and s, respectively. The p-value comes from performing a Student's t-test of the null hypothesis that data are independent random samples from normal distributions with equal means and equal but unknown variances, against the alternative that the means are not equal. $\star$ indicates a rejection of the null hypothesis at 5% significance level.

10.5 Discussion

Quantitative assessment of registration accuracy in Figure 10.2 indicates that the automatic method is more consistent than the human observers. This is also seen in Table 10.1 where the rms error is tabulated. In general, most rms-errors are below 2 mm. Hard tissue landmarks #1b-15b are easier to locate than #16b-24b for both human observers and automatic method; except for landmarks #4b, #5b and #10b, which are inconsistent for the human observers. Hard tissue landmark #21b shows a large error when using the automatic method. This landmark is placed at the posterior fontanelle, i.e. at the intersection between the sagittal and the lambdoid sutures. Landmark #23b is also difficult to locate; this is a point in the cranial base. The human observers show least agreement at landmark #4b-5b, #10b and #23b-24b. For soft tissue landmarks, landmarks #1s-12s and #17s are easy to locate for both human observers and the automatic method. Problems with landmarking might arise from lack of a clearly defined location e.g. no well-defined edge. According to Rohlfing [158], a (dense) set of landmarks is the ideal way of detecting incorrect registrations.

Usually a gold-standard requires more than two observers, as we know inter-observer variation exists. Ten or more human observers would have been appropriate for landmark validation as well as more landmarks. However, the task of landmarking is tedious and time consuming so we chose to support the analysis by incorporating point-to-surface errors. As seen in Table 10.2 and Figures 10.3 and 10.4, the registration errors are relatively small everywhere in the head/skull. We do, however, note that the fontanelles, cleft area, parts of the neck and ears are more difficult to match. Four of the patients have a relatively high point-to-surface error compared to the rest. The errors in three of these seem to be a result of mismatches in the area of the intubation device, while the errors for the last patient is related to the warped atlas being on the inside of the original surface.

Figure 10.5 shows the largest amount of variability in the chin/mandible area, neck area, eyebrow, back of the head and the cleft area. The chin/mandible deviations relate to differences in mouth opening; i.e. open versus closed mouth. The neck area is related to different head postures as the heads are aligned w.r.t. the voxel intensities. Differences in the forehead and back of the head are due to (deformational) plagiocephaly. 9 of the 17 children have plagiocephaly and three of these have right-sided plagiocephaly; one being a severe case. Deviations in the cleft area relates to width and height of the cleft; and for soft tissue open mouth and position of the lips also contributes to larger deviations.

Point-to-surface validation revealed small registration errors, but the fontanelles, cleft area, parts of the neck and ears were the most difficult to match. Problems

with intubation devices were responsible for large errors for three out of four bad cases, while the fourth case was related to thresholding. Table 10.3 shows a consistency in cleft width for the three measures on soft tissue. The bone tissue cleft width is wider due to the anatomy of the cleft. The cleft width using landmarks #15s-16s was rejected as having same means for manual and automatically placed landmarks. Looking at Figure 10.2 reveals that landmark #16s is a difficult landmark to locate for both human observers and the automatic method. Cleft height is consistent for both soft and bone tissue. The nose angle had no corresponding bone tissue measure, but was generated to check the consistency between manual and automatic annotation. Nose areas were as expected also similar for both soft and bone tissue.

The method, being fully automatic, has potentially a multitude of applications. Firstly, it is able to present detailed information about the mean and variation of shape in the pre-surgical UCLP population. The richness of data compared to previous X-ray cephalometric studies (e.g. those using the 3-projection cephalometric X-ray method developed by Hermann et al. [81]) opens up for investigations asking questions about the cause of UCLP and other types of cleft lip and/or palate. Is the malformation in cleft lip and palate spatially isolated to the cleft region, and what are the anatomical differences between various types of CLP? Moreover, the proposed automatic method would be interesting to test on other volumetric modalities, like MRI and CBCT, in a similar setup.

10.6 Conclusion

The landmark-based validation analysis revealed that the automatic approach performed on equal terms with the observers for the children with UCLP. The statistical tests concluded that the automatic method based on non-rigid registration was as accurate as the human observers'.

Anatomical variability in the group of infants with UCLP was determined and expressed as the standard deviation of the distances from all individuals and the mean. Variation in the neck region was found to relate to registration problems, while the deviations in the chin area, forehead, back of the head and cleft area were related to true anatomical variability e.g. plagiocephaly, width and height of the cleft, open/closed mouth.

Landmark-based measures were found to be consistent for the manual and automatic method, except in the case of a troublesome landmark that was hard to place for the human observers. Consistency was also found between soft and bone tissue.

10.7 Acknowledgment

For all image registrations the Image Registration Toolkit was used under license from Ixico Ltd. For visualizations of CT images and surfaces landmarker (http://www.labd.odont.ku.dk/landmarker/) [48] was used.

10.8 Appendix

10.8.1 Landmark Definitions

Bone tissue		**Soft tissue**	
Landmark	Definition	Landmark	Definition
1	Nasion	1	Nasion
2	Outer right eye corner	2	Outer right eye corner
3	Outer left eye corner	3	Outer left eye corner
4	Inner right eye corner	4	Inner right eye corner
5	Inner left eye corner	5	Inner left eye corner
6	Nose tip	6	Nose tip
7	Right nose wing	7	Right nose wing
8	Left nose wing	8	Left nose wing
9	Lowest part of cleft (right)	9	Septum
10	Lowest part of cleft (left)	10	Right lower cleft
11	Left coronoid process	11	Left lower cleft
12	Right coronoid process	12	Chin
13	Chin	13	Right ear
14	Right zygoma	14	Left ear
15	Left zygoma	15	Right upper cleft
16	Midpoint synchondrosis	16	Left upper cleft
17	Foramen magnum front	17	Right mid cleft
18	Foramen magnum back	18	Left mid cleft
19	Right ear canal	19	Right mouth corner
20	Left ear canal	20	Left mouth corner
21	Posterior fontanelle		
22	Sella turcica center		
23	Sella turcica left		
24	Sella turcica right		

Table 10.4: Anatomical interpretation of landmarks.

10.8.2 Comparing the error of an automatic method with the error of an observer in the absence of a gold standard

It is often desirable to be able to compare an automatic or semi-automatic method with the human error (the human error is assumed to be the truth). In many cases this is complicated by the fact that one does not know the truth, i.e. no gold-standard exists. Probably the most commonly cited reference for this and a similar kind of situation is that of Bland and Altman [22].

Two human observers placed points on a 3D structure. These points are to be compared to the placement of the same points by an automatic method. In such a setting it is possible to estimate the observer error (variance) and the error (variance) of the automatic method.

For simplicity assume the one-dimensional case where we ask each of the observers to mark the position of a point x. We assume the correct - but unknown - position is μ. We further assume that each human observer has his own bias and error variance and that they are independent of each other. Similarly the automatic method is assumed to have its own bias and error variance. We also assume independence between the automatic method and the observers[1]. Finally, the assumption of normality is useful in order to set up formal statistical tests. The above can be written

$$X_1 \in N\left(\mu_1, \sigma_1^2\right), \quad X_2 \in N\left(\mu_2, \sigma_2^2\right), \quad X_A \in N\left(\mu_A, \sigma_A^2\right), \tag{10.12}$$

where X_1, X_2 and X_A are independent and denote observer 1, observer 2 and the automatic method, respectively. At least the following two quantities are considered important:

1. Comparison of the two observers shown as the difference D between them: $D_{1,2} = X_1 - X_2$.

2. Comparison of the automatic method with the average of the observers: $D_{A,12} = X_A - \frac{(X_1+X_2)}{2}$

[1] In many cases the latter assumption is questionable, since the automatic method is often trained using the data from the observers with whom we wish to compare. However, we assume it is valid if we use cross-validation. Alternatively, the automatic method could be trained against a third observer.

It can easily be shown that these quantities are distributed as

$$D_{1,2} \in N\left(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2\right),$$
$$D_{A,12} \in N\left(\mu_A - \frac{\mu_1 + \mu_2}{2}, \sigma_A^2 + \frac{\sigma_1^2 + \sigma_2^2}{4}\right), \tag{10.13}$$

where $D_{1,2}$ and $D_{A,12}$ are independent. From the above we note that it is possible to test the differences in bias $\left(H_{01} : \mu_1 = \mu_2 \text{ and } H_{02} : \mu_A = \frac{\mu_1+\mu_2}{2}\right)$ using t-tests. Furthermore, if we introduce the average human observer variance as $\sigma_H^2 = \frac{\sigma_1^2+\sigma_2^2}{2}$, we get

$$D_{1,2} \in N\left(\mu_1 - \mu_2, 2\sigma_H^2\right),$$
$$D_{A,12} \in N\left(\mu_A - \frac{\mu_1 + \mu_2}{2}, \sigma_A^2 + \frac{\sigma_H^2}{2}\right), \tag{10.14}$$

where $D_{1,2}$ and $D_{A,12}$ are independent. From this we see that with no knowledge of the correct position (no gold-standard), the average observer variance can be estimated as half the variance of the differences between the observer positions. Furthermore, the automatic method variance can be estimated by subtracting the estimated observer variance from the empirical variance of the differences between the automatic method and the average of the two observers. Finally, we note that the two types of differences are independent and thus uncorrelated. A formal F-test of the hypothesis that e.g. the variance of the automatic method is equal to that of an observer $\left(H_0 = \sigma_A = \sigma_H^2\right)$ can now be performed by adjusting the respective empirical variances by factors $\frac{1}{2}$ and $\frac{2}{3}$.

The adjustment can also be performed directly on the differences using the factors $\frac{1}{\sqrt{2}}$ and $\sqrt{\frac{2}{3}}$ respectively. This is useful for plotting purposes since plots of the two types of differences can be compared more easily - especially when we can assume no bias (or at least that they are equal).

The above was derived for 1D. Similar derivations can be done for 3D. However, if positioning-error in the x, y, and z-dimension can be assumed independent of each other and with the same variance, then the variance estimates can be pooled. This is equivalent to having three times the number of observations.

Chapter 11

Dealing with Difficult Deformations: Construction of a Knowledge-based Deformation Atlas

Signe S. Thorup[†], Tron A. Darvann[‡,⋆], Nuno V. Hermann[‡,♯], Per Larsen[‡], H. Ólafsdóttir[†], Rasmus R. Paulsen[†], Alex A. Kane[□], Dan Govier[□], L.-J. Lo[△], Sven Kreiborg[‡,♯,◇], Rasmus Larsen[†]

[†)] DTU Informatics, Technical University of Denmark, Lyngby, Denmark
[‡)] 3D Craniofacial Image Research Laboratory, Copenhagen, Denmark
[⋆)] Dept. of Oral and Maxillofacial Surgery, Copenhagen University Hospital, Copenhagen, Denmark
[♯)] Pediatric Dentistry and Clinical Genetics, University of Copenhagen, Denmark
[◇)] Dept. of Clinical Genetics, Copenhagen University Hospital, Copenhagen, Denmark
[□)] Dept. of Plastic and Reconstructive Surgery, Washington University School of Medicine, St. Louis, MO, USA.
[△)] Dept. of Plastic and Reconstructive Surgery, Chang Gung Memorial Hospital and Chang Gung University, Taipei, Taiwan.

ABSTRACT

Twenty-three Taiwanese infants with unilateral cleft lip and palate (UCLP) were CT-scanned before lip repair at the age of 3 months, and again after lip repair at the age of 12 months. In order to evaluate the surgical result, detailed point correspondence between pre- and post-surgical images was needed. We have previously demonstrated that non-rigid registration using B-splines is able to provide automated determination of point correspondences in populations of infants without clefts[1]. However, this type of registration fails when applied to the task of determining the complex deformation from before to after lip closure in infants with UCLP. The purpose of the present work was to show that use of prior information about typical deformations due to lip closure, through the construction of a knowledge-based atlas of deformations, could overcome the problem. Initially, mean volumes (atlases) for the pre- and post-surgical populations, respectively, were automatically constructed by non-rigid registration. An expert placed corresponding landmarks in the cleft area in the two atlases; this provided prior information used to build a knowledge-based deformation atlas. We model the change from pre- to post-surgery using a thin-plate spline warping. The registration results are convincing and represent a first move towards an automatic registration method for dealing with difficult deformations due to this type of surgery.

New or breakthrough work to be presented: The method provides a simple way of dealing with complex morphological changes using knowledge of typical deformations.

Index Terms— Non rigid image registration, computational atlas, difficult deformations, computed tomography, cleft lip and palate, change due to surgery.

11.1 Introduction

A common purpose of non-rigid image registration is to determine the transformation that would bring two images into detailed spatial correspondence. The correspondences provide a deformation field that may be used in order to study the differences between the images, for example representing differences between two subjects or temporal change of the same subject.

A prerequisite for a successful registration is that the images involved are similar: corresponding structures must exist in the two images, and the structures must

[1]Errata: Reference to [132]

be located within the spatial catchment region of the particular registration algorithm being used. If subjecting the registration algorithm to the task of registering increasingly different images, it will eventually fail. When studying temporal change, for example due to growth, the time separation between images needs to be tuned to the speed of the imaged process. In medical imaging, for example of infants with craniofacial malformations, the changes occurring between images are often large due to restrictions on acquisition frequency, radiation considerations or impracticalities of acquiring images intra-operatively. The study of changes due to surgery may be ill-posed if images become (too) different because of re-arrangement of, or removal/addition of tissue, leading to non-diffeomorphic deformations or removal of correspondences. Some types of surgery lead to preservation of correspondences, but the resulting deformation being locally too large for the registration to be successful. In such cases, a remedy could be to assist the registration by manual input from the user. For example, the user could point out corresponding anatomical landmarks forming a basis for an interpolation, e.g. by use of thin-plate splines (TPS) [75], of a transformation I that would serve as an initialization for the registration algorithm. In this case, the true transformation F to bring the two images into registration would be approximated by a composition of I and the transformation T subsequently determined by the non-rigid registration algorithm: $x' = F(x) = T(I(x))$. However, such a strategy would lead to a depreciation of the registration method from being automatic to becoming semi-automatic.

An alternative would be to automate the initialization step, provided that the transformation I could be predicted. In some types of surgery this may turn out to be possible. The present contribution demonstrates the feasibility of such an approach in one particular type of surgery: the so-called lip adhesion carried out in order to close the cleft of the lip in infants with cleft lip and/or palate (CLP).

CLP is the most common congenital craniofacial malformation with an incidence of about 1:500 live births [80]. It is caused by incomplete fusion of the maxillary and palatal processes early in the embryonic life [116]. CLP can be uni- (UCLP) or bilateral (BCLP) and may involve soft tissues (lip and/or soft tissues in the alveolus and/or palate) and/or bone (alveolus and/or hard palate).

CLP is typically treated by surgical closure of the lip and palate during the first two years of life. The first surgical procedure takes place at about 3-5 months of age and aims at closing the cleft of the lip. Many surgical approaches exist, but the so-called lip adhesion [116] is a common way of closing the cleft by stretching the soft tissue of the lip across the cleft. In the case of an additional clefting of the bone (alveolus and/or hard palate), the forces from the lip tissue stretching will also lead to a slight narrowing of the bony cleft during the time from lip adhesion to the time of surgical closure of the palate which takes place approximately a year later. In the present study, images were acquired of infants

with UCLP just prior to lip adhesion (at 3 months of age) and again just prior to palatal closure (at 12 months of age)[2]. In addition to the change due to growth, there is, thus, a deformation due to the lip adhesion that is very large in the soft tissue of the lip, while just adjacent to this large deformation, a smaller deformation has taken place in the region of the bony cleft.

An attempt to quantify the deformation field from before to after surgery by use of the non-rigid image registration algorithm developed by Rueckert and coworkers [163, 161, 170, 162], henceforth referred to as Rueckert's algorithm, using B-splines and normalized mutual information (NMI) [187], failed in the proximity of the cleft. But was seen to give reasonable results after an initialization in terms of a TPS deformation based on a few manually placed landmarks. In contrast, Rueckert's algorithm was seen to be successful when applied to the task of registering pairs of before-surgery images, or pairs of after-surgery images, in addition to similar registration tasks demonstrated in Rueckert [163, 161].

A hypothesis that the initial transformation I could be predicted from a sample population of infants was formulated after a few observations or assumptions had been made:

1. The pre-surgical shape of the face, including the cleft, is a naturally occurring phenomenon and may, in a population of infants with the same cleft type (e.g. UCLP), be modeled in terms of a normal distribution, as demonstrated e.g. in Darvann [47].
2. Lip adhesion, closing the gap between the lip segments, may be regarded as a process where tissue is neither re-arranged, nor added or removed, and where detailed point correspondence from before to after surgery exists, representing a diffeomorphic deformation field.
3. There is a correlation between the pre-surgical shape of the cleft region and the deformation of that region due to the surgery. The surgery may thus, as a consequence of (1), also be modeled in terms of a normal distribution.
4. In particular, (3) holds true if the surgery is carried out with the same protocol for all infants, and the surgery, as well as imaging, takes place at similar ages for all infants.

The hypothesis states that, in a population of infants with UCLP, the initial transformation I may, due to (3), be predicted from the deformation (representing a transformation A) of the individual pre-surgery image from the mean pre-surgery image (the pre-surgery atlas). A may be determined automatically by use of Rueckert's registration algorithm.

[2]Errata: Prior to palate closure i.e. after lip repair.

In particular, a related hypothesis states that a sufficient initial transformation I consists of the mean transformation for all individuals in the population; i.e. the transformation that brings the pre-surgery atlas into correspondence with the post-surgery atlas.

It is the latter hypothesis that we test and validate in the present article.

11.2 Materials

11.2.1 Data

The data consisted of CT scans of 23 Taiwanese infants with Unilateral Cleft Lip and Palate (UCLP). All data were obtained in a clinical context at Chang Gung Memorial Hospital in Taipei, Taiwan. The use of CT was approved by a local ethics committee at the hospital. The infants were scanned before lip repair at the age of 3 months, and again after lip repair at the age of 12 months. Image pixel resolution varied from 0.5 to 1.0 mm, and slice distance varied from 1.0 to 2.0 mm[3]. Figure 11.1a and 11.1b show soft tissue skin surfaces extracted from CT-scans of one of the individuals, using the Marching Cubes algorithm [114].

11.3 Methods

11.3.1 Data preprocessing

Prior to analysis, images were converted to have isotropic voxel resolution of $0.5 \times 0.5 \times 0.5$ mm. UCLP is either right- or left-sided, and therefore, in order to increase the sample size[4], all CT-scans of infants with right-sided clefts were mirrored to become left-sided as a pre-processing step prior to analysis.

[3]Errata: In-slice spatial pixel resolution ranged from 0.35 to 0.49 mm, and slice distance varied from 1.5 to 3.0 mm. The scans were reformatted in Taiwan to $340 \times 340 \times 400$ with equal resolution of 0.5 mm.

[4]Errata: Data size.

11.3.2 Image registration

The goal of image registration is to warp one image, the source, into the coordinate system of another image, the target, using a transformation T. A basic image registration algorithm requires the following: a transformation model, T, a measure of image similarity, and an optimization method to optimize the similarity measure with respect to the transformation parameters.In this study, two different transformation types were used, an affine transformation, and a non-rigid deformation based on B-splines [163, 161]. The first captures global, linear differences between the images while the latter handles local, nonlinear differences. In both cases, normalized mutual information (NMI) [187] was used as a similarity measure and gradient descent optimization was applied.

11.3.2.1 Affine registration

Affine registration applies an affine transformation to map the source image into the target image. An affine transformation is a linear transformation

$$T_{affine}(x, y, z) = Qx + t \tag{11.1}$$

where, in 3D, x is a vector containing the 3D point coordinates (x, y, z), t is a vector containing the three translation parameters and Q is a 3×3 matrix of parameters, including rotation, scaling, skewing and shearing [184]. For all registrations including atlas construction 9 transformation parameters (translation, rotation and anisotropic scaling) were used.

11.3.2.2 Non-rigid registration based on B-splines

To obtain a more accurate registration focusing on local differences, nonlinear transformations are required. A widely used method for this purpose is the non-rigid registration algorithm using B-spline-based free-form deformations (FFDs) [163, 161]. In this case, a composition of a global and a local transformation,

$$T(x, y, z) = T_{global}(x, y, z) + T_{local}(x, y, z) \tag{11.2}$$

is applied. The global model corresponds to the affine transformation. In 3D, the local transformation model, the FFD is defined by an $n_x \times n_y \times n_z$ mesh of control points Φ with spacing $(\delta x, \delta y, \delta z)$. The underlying image is deformed by manipulating the mesh of control points. The FFD model can be written as the

tensor product of the one-dimensional (1D) cubic B-splines:

$$T_{local}(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\phi_{i+l,j+m,k+n} \tag{11.3}$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $k = \lfloor z/n_z \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$ and $w = z/n_z - \lfloor z/n_z \rfloor$. $B0$ through $B3$ represent the basis-functions of the B-spline:

$$\begin{aligned} B_0(u) &= (1-u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6 \end{aligned}$$

The transformation creates a dense deformation vector field which can be assessed at any point in the image.

11.3.2.3 Atlas construction

Two computational, deformable atlases were constructed from the sample of infant CT scans; one from 18 individuals before surgery and one from the same 18 individuals after surgery. The atlases were constructed in an iterative manner using non-rigid registration (a composition of an affine registration and B-spline-based non-rigid registration). The atlases were built according to the procedure listed in Algorithm 1 [74].

Algorithm 1: Automatic atlas construction	
1	Atlas = a selected reference image of an infant from the group (before or after surgery)
2	**Repeat:**
3	Register all images from the given group to atlas.
4	Atlas = Intensity average of all registered images.
5	**until** atlas stops changing
6	Register atlas to all images from the given group.
7	Transform atlas by the average transformation obtained in step 6.

In other words, we first select a reference image to be the basis for the atlas construction within each of the two groups. Subsequently, the rest of the images were registered to the reference image in order to create an intensity average atlas. In line 5 of Algorithm 1, the root-mean-square (rms) error between the voxel intensities of the current atlas and the previous atlas was calculated and an appropriate threshold value was chosen in order to define the state where the atlas stops changing. Lines 6 and 7 of Algorithm 1 are intended to reduce the bias in shape towards the chosen reference subject, as previously demonstrated with good results [133, 192]. The non-rigid registration was carried out in a coarse-to-fine manner with successive control point spacings of 8 mm, 4 mm and 2 mm. The initial spacing represented the larger scale non-rigid deformations with the subsequent spacings used to capture increasingly finer detail. The FFD resulting from optimizing each control point spacing was used as an initial estimate for the next.

11.4 Experimental results

11.4.1 Proposed method

To overcome the problems visualized in Figure 11.1c[5], it is proposed to bridge the large change between the pre- and post-surgery-atlases by an initial non-rigid deformation field I determined by thin-plate-splines (TPS) applied to manually placed landmarks. Anatomical landmarks were placed in the eye, nose and mouth regions: 22 on the skin surface, 24 on the skull bone surface. Additionally,

[5]Errata: The cleft persists to be there even though it is supposed to be closed.

23 pseudo-landmarks were placed in order to control the TPS. The method is motivated by the observation that a point distribution model containing landmarks from both ages is distinctly bi-modal [47], making it probable that I will be a good initial guess for the deformation field for any UCLP lip surgery. The proposed algorithm is described in the following.

Algorithm 2: Automatic determination of the deformation field due to pre- to post lip closure surgery in an infant with UCLP	
1	Register the pre-surgery image to the pre-surgery atlas using an affine transformation.
2	Register the aligned pre-surgery image (result of (1)) to the pre-surgery atlas using non-rigid image registration. The result is an image of the individual after it has been brought into best correspondence with the atlas, using a transformation A.
3	Register the result from (2) to the post-surgery atlas using a landmark-based TPS transformation I followed by non-rigid image registration to make the TPS-deformed image into best correspondence with the post-surgery atlas, using a transformation T_{TPS}.
4	Register the result from (3) to the image of the individual after surgery using non-rigid image registration, using a transformation B.
5	The final deformation field between the pre- and post- surgery images for the individual may now be expressed in terms of a transformation being a composition of the four computed transformations A, I, T_{TPS} and B.

Note that the method is fully automatic, since manual landmarks need to be placed only in the atlases, predicting F for an (unseen) individual. A schematic drawing of Algorithm 2 is shown in Figure 11.2.

11.4.2 Validation

Validation was carried out using surfaces extracted from the CT volumes by means of the Marching Cubes algorithm [114]. The distance from each point of a registered surface to the closest point of the original surface was calculated, representing registration error [48]. The errors were visualized on bone surfaces as well as on facial skin soft tissue surfaces for steps 2-4 of Algorithm 2. The validation was carried out using five CT-scans that were not included in the

atlas construction. Figure 11.1c and d provides a visual comparison between the results of registrations using the original, direct registration, and the proposed algorithm where the transformation T_{TPS} contains the predicted transformation I.

Figure 11.3 presents the validation results for a pre- to post-surgery registration in a single individual. The three left-most columns in the Figure correspond to the three transformations A, T_{TPS}, B of Algorithm 2, respectively, while the fourth column corresponds to the direct registration between the pre- and post-surgery images. The first column of Figure 11.3 shows surface reconstructions of the result of deforming the image of the pre-surgery individual using the transformation A (step 1 and 2 in Algorithm 2) to bring it into best correspondence with the pre-surgery atlas, color coded according to surface-to-surface error. The second column shows surface reconstructions of the combined result of the transformations A and T_{TPS}, color coded according to the error of the registration corresponding to the transformation T_{TPS} (step 3 in Algorithm 2). Third column shows surface reconstructions of the combined result of the transformations A, T_{TPS} and B, color coded according to the error of the registration corresponding to the transformation B (step 4 in Algorithm 2). The right-most column shows the result of the direct registration between pre- and post-surgery images of the individual, color coded according to the error of that registration. To summarize the error analysis, Table 11.1 presents the mean (for n=5 subjects not included in the atlases; atlases containing 18 subjects) error distances for bone surface and skin soft tissue surface for each of the four registrations, respectively (corresponding to transformations A, T_{TPS}, B and F, respectively). To further analyze the errors, the spatial error distribution is shown in Figure 11.4 for the bone surface and skin soft tissue surface for each of the deformations.

11.5 Discussion

Qualitative evaluation of the relative performance of the registration algorithms (Rueckert's algorithm versus Rueckert's algorithm including the TPS initialization), carried out by visual inspection of registration results as those shown in Figure 11.1c and d, demonstrated that the latter method performed convincingly well, and seemed to be able to correctly quantify the difficult deformation taking place from before surgery to after surgery. The quantitative validation of the registration in the part of the volume corresponding to bone surface shows that most regions are accurately matched with registration error between 0 and 1 mm. Some large errors occur in the neck region, on the edges of the fontanelle, as well as on erupting teeth. The latter may be caused by the presence of different stages of tooth eruption in the different individuals. Overall, the results of registration

using the proposed method are very promising in terms of accurate registration in areas corresponding to the bone surface. For the validation in the soft tissue skin, some problems are seen to occur in the ear and neck region. The outer soft-tissue ears[6] are complicated structures and are therefore difficult to match. However, the investigation of morphology of the outer soft-tissue ears[7] is not our primary goal. Large registration errors also occur in the neck and shoulder region, since these regions are present in only some of the images. In Figure 11.3 it is also observed that the errors for the registration corresponding to transformation B are, in most regions, larger than for the transformations A and T_{TPS} when assessed on the soft tissue skin surface. For example, there are errors in excess of 1 mm in the forehead and along the bridge of the nose. The reason for this is not fully understood, but it is probably related to the accumulating errors coming from composing the transformations A, T_{TPS} and B or related to the smooth atlas being registered to an individual with distinctly marked edges. However, in the important region in the proximity of the cleft and mouth, the errors are below 1 mm, which is very good. Comparing the output from registration corresponding to transformation F to the output of B, we notice that the former in general performs very well except in the cleft area, where errors larger than 3 mm are seen to occur. Table 11.1 confirms that registration errors are very low on average when assessed on the bone surface. On the soft tissue skin surface, the errors are also low, except for registration corresponding to transformation B, again probably due to accumulation of errors. Omitting the ears from the validation would improve the registration errors corresponding to B to some degree. Comparing the transformations B and F we again find that F has the smallest error and with the least variation when assessed on soft tissue skin surface. However, the errors for bone are almost the same for the two methods, and the standard deviation for B is actually lower than the one for F. The spatial error distribution in Figure 11.4 summarizes the error analysis. It shows that the error mean is below 1 mm for most of the surface, both for bone and soft tissue skin. The majority of the errors have standard deviation of around 0.5 mm for both bone and skin. About 4 % of the total number of bone surface triangles for registration corresponding to the transformation B has errors larger than 3 mm. When interpreting the validation based on surface-to-surface distances it is important to realize that these small distances do not necessarily imply that the registration was successful in the anatomical sense. This is particularly true for large or difficult deformations like the ones occurring in the cleft area due to the lip adhesion. From looking at Figure 11.3, results of transformations T_{TPS} and F, it is evident that the validation method under-estimates the errors. Therefore, a landmark based validation (see e.g. Thorup [192]) is planned for the future. Still, the combination of the qualitative inspection and the surface-to-surface distance based quantitative validation gives

[6]Errata: The soft tissue of the outer ear.
[7]Errata: The soft tissue of the outer ear.

clear evidence that the method we propose is feasible and provides accurate results.

11.6 Conclusion

We have presented an extension to Rueckert's non-rigid registration algorithm [163, 161, 170] making it able to deal with difficult deformations due to primary lip surgery in infants with UCLP. The extension consists of providing an initialization in terms of a non-rigid transformation created by TPS applied to a number of landmarks. Due to properties of the deformation, the method is fully automatic and predicts a set of landmarks and corresponding initial TPS transformation. In the current article we showed that the method provided good registration results even if the transformation between atlases was used as initialization. The successful registration obtained with the proposed method makes it possible to automatically and objectively establish detailed volumetric point correspondence between images representing pre- and post-surgical anatomy, both in the proximity of the cleft as well as in the rest of the craniofacial region. The new method makes it possible to study changes due to growth and surgery in infants with CLP.

11.7 Acknowledgement

For all image registrations the Image Registration Toolkit was used under license from Ixico Ltd. For visualizations of CT images and surfaces landmarker (http://www.labd.odont.ku.dk/landmarker/) [48] was used.

11.8 Tables and Figures

Mean errors in mm	**Transformation** A	**Transformation** $T_{\mathbf{TPS}}$	**Transformation** B	**Transformation** F
Soft tissue skin	0.31 ± 0.26	0.38 ± 0.72	0.98 ± 1.28	0.38 ± 0.50
Bone surface	0.34 ± 0.30	0.26 ± 0.49	0.38 ± 0.52	0.36 ± 0.88

Table 11.1: Mean surface to surface registration errors $\pm$ one standard deviation [mm] for each of the registrations listed in Algorithm 2, step 2-5.

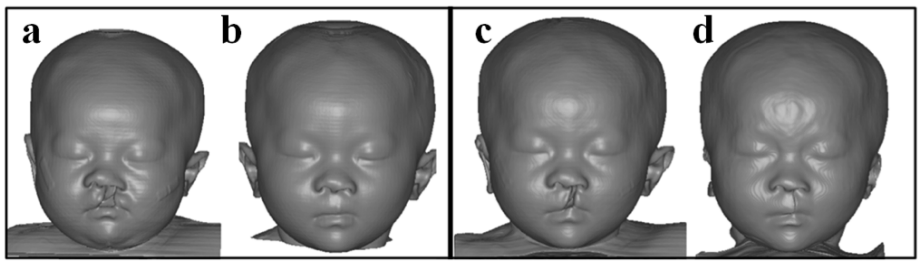


Figure 11.1: Example infant soft tissue skin surfaces (left box) and results of non-rigid registration (right box). **a**: Infant before lip adhesion at 3 months of age. **b**: Same infant 9 months after lip adhesion. **c**: Result of applying non-rigid registration (Rueckert's algorithm) to the pre-surgery image **a** to bring it into correspondence with the post-surgery image **b**. Note the excellent result except in the proximity of the cleft. **d**: Result of applying non-rigid registration (proposed indirect algorithm) to the same task as in **b**. Note the improved result in the cleft region.

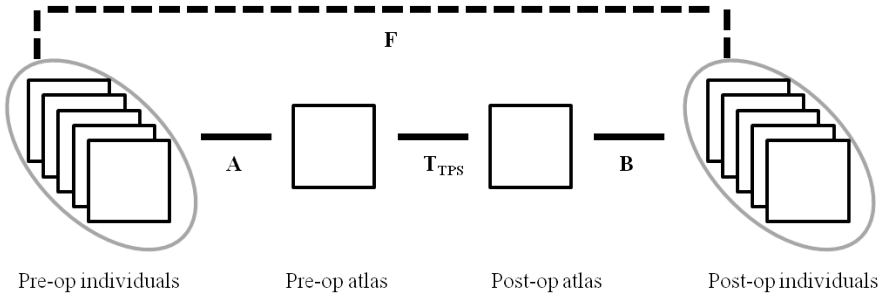


Figure 11.2: Schematic drawing of Algorithm 2. The proposed method achieves the difficult registration between pre- and post-surgery images, corresponding to the transformation F, as a composition of the transformations A, T_{TPS} (including the initialization I) and B.

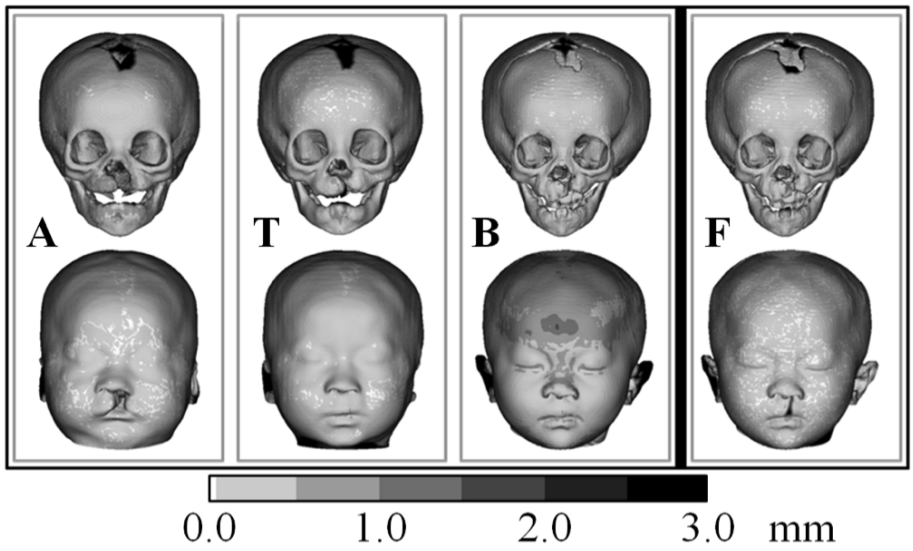


Figure 11.3: Surface to surface registration errors for the bone (top) and soft tissue skin surface (bottom), for one UCLP subject. From left to right, the soft and bone tissue errors for registrations defining the transformations A, T_{TPS}, B, and F respectively, are shown. The colors denote errors in mm according to the color bar, where white denotes no errors, and black denotes errors of 2.5 mm and above.

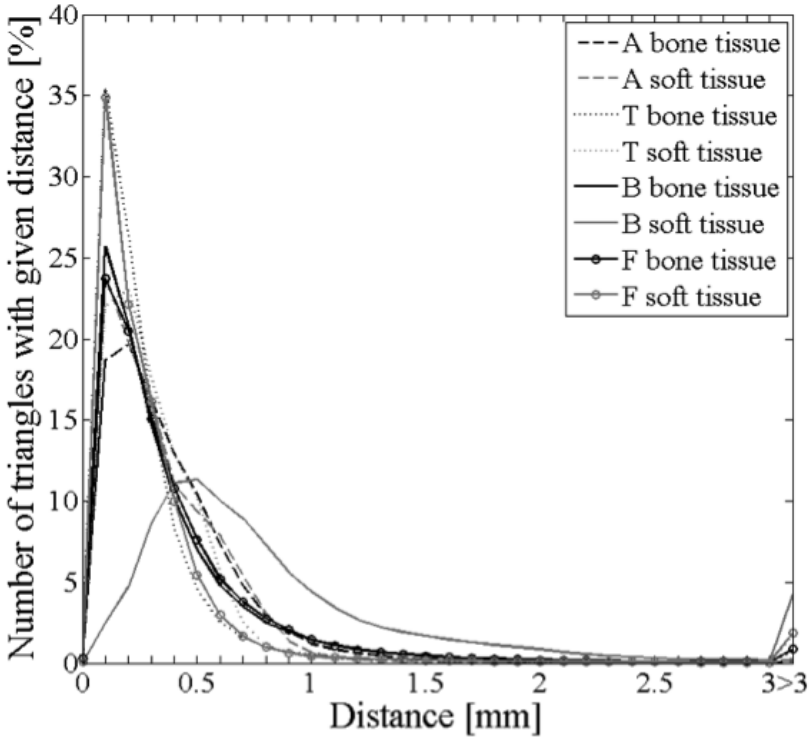


Figure 11.4: Spatial distribution of mean registration errors. % of surface points (center of triangles) with the given error are shown for each type of surface (bone, soft tissue skin) and for each of the registrations indicated in Table 11.1

Chapter 12

Quantification of Craniofacial Growth in Mice with Craniofacial Dysmorphology Caused by the Crouzon Mutation Fgfr2^{C342Y}

Signe S. Thorup[†], Tron A. Darvann[‡,⋆], Hildur Ólafsdóttir[†], Nuno V. Hermann[‡,♯], Rasmus R. Paulsen[†], Per Larsen[‡], Chad A. Perlyn[□], Rasmus Larsen[†], Sven Kreiborg[‡,♯,◇]

[†)] DTU Informatics, Technical University of Denmark, Lyngby, Denmark
[‡)] 3D Craniofacial Image Research Laboratory, Copenhagen, Denmark
[⋆)] Dept. of Oral and Maxillofacial Surgery, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark
[♯)] Pediatric Dentistry and Clinical Genetics, School of Dentistry, University of Copenhagen, Copenhagen, Denmark
[□)] Division of Plastic Surgery, Miami Children's Hospital, Miami, FL, USA
[◇)] Dept. of Clinical Genetics, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

ABSTRACT

Crouzon syndrome is characterized by premature fusion of craniofacial sutures and synchondroses of the cranial base, which can lead to abnormal head shape and disturbed facial growth. In 2004, the first mouse model, equivalent to the most common human Crouzon/Pfeiffer syndrome mutation, was introduced. Using the Crouzon mouse model, our goal was to quantify, analyze and compare patterns of normal and abnormal craniofacial growth from a cross-sectional set of micro-CT volumes of normal and Crouzon mice.

Local changes in morphology were modeled using non-rigid image registration, providing growth patterns automatically and with greater detail than methods employing manually placed landmarks. Validation of the modeling was conducted by a comparison to manually placed landmarks, and the growth patterns were evaluated by a group of clinical experts.

For growth analysis, different approaches were employed to visualize and estimate significant areas of growth and growth differences between Crouzon mice and wild-type mice: construction of animations, application of statistical methods (local False Discovery Rate (FDR) analysis, Multivariate Analysis of Variance (MANOVA)), and landmark-based skull measurements.

Most registration errors, comparing automatic to manually placed landmarks, were below 0.5 mm, meaning that the accuracy was within the range of previously estimated errors for a similar dataset. Furthermore, the growth pattern was quantified and found realistic in terms of existing knowledge on humans. With the automatically placed landmarks, we were able to estimate the growth rate for Crouzon and normal mice. On average, skull length and nose length expanded almost twice as fast for normal mice compared to Crouzon mice. While on average; skull height, skull width and interorbital distance expanded twice as fast for Crouzon mice compared to normal mice.

Index Terms— Craniofacial growth, Crouzon syndrome, mouse model, non-rigid image registration, computational atlas, craniofacial mouse atlas, morphology changes, cross-sectional data

12.1 Introduction

Crouzon syndrome was first described about a century ago when calvarial deformities, facial anomalies, and abnormal protrusion of the eyeballs were

reported in a mother and her son [43]. Later, the condition was found to be characterized by premature fusion of the cranial sutures (craniosynostosis), orbital deformity, maxillary hypoplasia, beaked nose, crowding of teeth, and high arched or cleft palate. In the vicinity of the fused suture, growth is inhibited and the brain cannot expand normally. Consequently, the brain expands where it can, depending on the number of sutures involved, and this shapes the skull as the order and progression rate of the suture fusion allows. For example, if the craniosynostosis is taking place early, the dysmorphology would be expected to be more pronounced [102, 39, 150].

Heterozygous mutations in the gene encoding fibroblast growth factor receptor type 2 ($Fgfr2$) have been found to be responsible for the Crouzon syndrome [148]. The mutations lead to increased osteoblast differentiation [128], which consequently increases the probability of premature fusion of sutures and synchondroses.

In children with Crouzon syndrome, the coronal sutures are most often fused at birth, which influences volume change for the brain and leads to cranial dysmorphology. Postnatal progressive fusion of sutures of the calvaria, cranial base, orbits and maxillary complex and the synchondroses of the cranial base are characteristic. Since the order and progression rate of suture fusion seem to be random, variation in craniofacial dysmorphology is high. For instance, uni-coronal synostosis is expected to induce more asymmetry than a bilateral synostosis.

Genetic alteration of the mouse genome has become a standard tool in the field of craniofacial developmental biology. Numerous mouse models for craniofacial anomalies now exist, each with a unique phenotype [196]. In 2004, a mouse model was created to study one mutation associated with Crouzon syndrome ($Fgfr2^{Cys342Tyr}$) [62]. In two previous studies, this model was analyzed using micro-CT scans of a group of unaffected wild-type mice and a group of mice with Crouzon syndrome. Both studies confirmed that the mouse model reflected the craniofacial deviations occurring in humans with Crouzon syndrome, thus validating many of the morphological traits [140, 133]. [140] used a comparison of linear measurements obtained manually on the mouse skulls; [133] automated the phenotyping of the adult 6-week-old mice and added more details to the study by applying image registration.

The primary goal of the present study was to quantify craniofacial growth and growth deviations in the Crouzon group relative to a control group. Using techniques from image analysis, this paper presents the automatic construction of a growth model directly from micro-CT data. The growth analysis should be able to visualize and quantify significant areas of growth and growth differences between Crouzon mice and wild-type mice. For this purpose we constructed

animations, applied statistical methods (local FDR analysis; MANOVA) and landmark-based skull measurements. Furthermore, we'll discuss the pros and cons of these four methods. Due to limited sample sizes and cross-sectional data, this study focuses on average craniofacial growth.

The paper is organized as follows: the methods section covers data acquisition followed by an introduction to image analysis methods and growth modeling. The results section provides experimental results by a qualitative and a quantitative validation of the automatic assessments as well as the analysis of the different growth outcomes in Crouzon mice and wild-type mice. A description of the MANOVA, used for assessing statistical significance of growth, is found in the Supplementary Material 12.4.

12.2 Main Body

12.2.1 Data

Micro-CT scans of 4-week-old mice (five specimens) and 6-week-old mice (ten specimens) with Crouzon syndrome ($Fgfr2^{C342Y/+}$) were compared to similar control groups of wild-type mice. The production of the $Fgfr2^{C342Y/+}$ mutant mouse (Crouzon mouse) was carried out as described by [62]. All procedures were in agreement with the United Kingdom Animals (Scientific Procedures) Act, guidelines of the Home Office, and regulations of the University of Oxford. Mutant mice of breeding age were determined by phenotype. Female $Fgfr2^{C342Y/+}$ mice were bred with males heterozygous for the same mutation.

For micro-CT scanning, the 15 wild-type and 15 $Fgfr2^{C342Y/+}$ specimens at 4 and 6 weeks of age (28 and 42 days), respectively, were euthanized by CO_2 asphyxiation and whole mount skeletal preparations were carried out. The mice were sealed in conical tubes and shipped to the micro-CT imaging facility at the University of Utah, Salt Lake City, USA. 3D volumes of the skull of size $720 \times 480 \times 480$ voxels were obtained at approximately $46\mu m \times 46\mu m \times 46\mu m$ resolution per voxel using a General Electric Medical Systems EVS-RS9 Micro-CT scanner.

At the age of 4 weeks, the coronal sutures were fused; there was a partial fusion of the lambdoid sutures; and the sagittal suture was partially separable. By the age of 6 weeks the sagittal suture was completely obliterated in 70% of the mice. Prior to image analysis, the neck part, hyoid bone and scanning artifacts were removed. Figure 12.1 gives an example of the imaging data appearance, after

surface extraction [114] from the micro-CT.

12.2.2 Methods

12.2.2.1 Image Registration

In the following, all operations explained are performed in three dimensions on 3D volumetric images. Because, technically, the volumes are 3D images, the terms image and volume are used interchangeably in the following.

A typical image registration algorithm that aims to match two images to obtain anatomical correspondence consists of three parts:

- A transformation model, T.
- A measure of image similarity.
- An optimization method to optimize the similarity measure with respect to the transformation parameters.

A transformation model is required to mathematically describe the relationship between corresponding anatomical points in the two images. The similarity measure is required to indicate how accurately the images match, and finally the optimization method is applied to optimize the similarity measure with respect to the parameters of the transformation model.

The non-rigid registration algorithm applied for this study is adopted from [163] and [170]. This algorithm uses a transformation model T, which is a combination of global, (T_{global}) and local (T_{local}) transformations,

$$T = T_{\text{global}} + T_{\text{local}}. \tag{12.1}$$

The global transformation model consists in our case of an affine transformation model (translation, rotation and scale) capturing global, linear differences between the images. Depending on the choice of parameters, scale may not be included. The more interesting local transformation model, providing the non-rigid part of the model, is described by a free-form deformation model (FFD). This is defined by a $px \times py \times pz$ mesh of control points c with spacing $(\delta x, \delta y, \delta z)$. The underlying image is then deformed by manipulating the mesh of

control points. The model is written as the tensor product of 1D cubic B-splines,

$$T_{local}(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)c_{i+l,j+m,k+n}. \tag{12.2}$$

$B0$ to $B3$ are the pre-defined basis functions of the B-spline, i, j and k are control point indices and u, v and w are (x, y, z) image coordinates translated into the mesh coordinates.

As a similarity measure, we have applied Normalized Mutual Information (NMI) as described by Studholme et. al [187] and carried out optimizations using a gradient descent approach as described by Rueckert et. al [163].

Computational atlases have many applications. In most cases, they are deformable, meaning that it is possible to deform them into the corresponding anatomy of any subject within a population. These properties allow automatic linear or volumetric measurements and segmentation of different structures and organs (e.g. [54, 115, 137, 44, 210]. Quantification of growth requires assessment of local deformations (growth). Often, the anatomy of a single, normal healthy subject is referred to as an atlas and used as a reference frame when estimating deviations associated with the pathology. A more unbiased way of defining such a reference frame is to use the average of a set of normal subjects. This can be a set of points delineating an anatomical structure averaged over a set of subjects [189]. Inclusion of more anatomical details results in shape-based and intensity-based atlases constructed from a set of images in 2D, 3D, and 4D [29, 38, 90, 142]. We chose to build atlases according to the unbiased non-rigid procedure described by [161].

12.2.2.2 Image Registration Parameters

Pre-processing of the data using the affine transformation was carried out as an initialization for the non-rigid image registration and to ensure that all mice faced the same way. Recalling that the transformation model T_{global} can account for size, atlases were built incorporating the scaling parameter. Growth modeling, on the other hand, is a matter of determining size changes, and thus scaling was left out. Following the global registration, non-rigid registration was carried out in a coarse-to-fine manner with successive control point spacings of 3 mm, 1.5 mm and 0.75 mm. The initial spacing represents the larger scale non-rigid deformations with the subsequent spacings used to capture increasingly finer

detail. The FFD resulting from optimizing the control point spacings was used as an initial estimate for the next.

In cases where the 6-week-old wild-type atlas was chosen as a common system of reference for comparing normal and abnormal growth, control point spacings of 6 mm, 3 mm, 1.5 mm and 0.75 mm were used. The control spacing of 6 mm was introduced to avoid complications in the registration due to the very different sizes and shapes of the Crouzon mice compared to the atlas.

12.2.2.3 Growth Modeling

In the present work, a growth model was built from micro-CT scans of mice; thus potentially modeling changes at a spatial scale occurring down to the image resolution of 46 μm voxel. The growth is modeled as the rate of change of the position (velocity or growth vector, $v = \frac{dp}{dt}$) of each anatomical structure. By applying non-rigid image registration, corresponding anatomical structures are localized and mapped in images of mice at different ages. The spatial resolution of the model is dependent on the non-rigid registration algorithm to be able to match small structures between the images. Some loss of spatial resolution is inherent in the method due to the use of an atlas built from a group of mice. Furthermore, since the mice were euthanized before micro-CT scanning, a spatial smoothing takes place due to the cross-sectional data i.e. comparing different mice with different local anatomical structures. As a consequence, the growth model created represents the average growth of the group of mice during the time span dt. The effective performance of the spatial registration and, consequently, the spatial resolution of the growth model, are checked by validation against manual annotations performed by an expert.

A growth model is a mathematical description of the change of a system. The purpose of the model is to investigate the properties of the system and potentially predict future states. A mouse skull is a complex system and a detailed growth model would need to take into account local variations due to osteoblast and osteoclast activity at a microscopic level. A more global growth model would be able to mimic the large scale properties of the growing mouse skull.

The global nature of the model, capturing the large-scale growth of the skull, matches the macroscopic level of detail used during the typical process of phenotyping based on skull shape. Recognizing a craniofacial syndrome typically involves visual features from the global skull shape; some features are obvious while others are subtle and, therefore, difficult to quantify. A successful growth model would be able to capture and quantify these features and their change over time.

For better visualization and validation of the model, surfaces were extracted from the micro-CT scans. A detailed point correspondence was needed to map the growth vectors (or deformation vectors) from the non-rigid volume registration to surfaces. The point correspondences were achieved using the procedure listed in Algorithm 1, Table 12.1. The algorithm creates a new surface, which shares the same vertex-number and vertex-order as the target mouse, but still resembles the source mouse.

The image registration process provides growth vectors at every voxel throughout the image. Since no soft tissues were included in the images, the growth model was built after extracting a bone surface using the marching cubes algorithm [114] for better visualization.

12.2.2.4 Registration Accuracy

Validation of the registrations is interesting as registration becomes more difficult the less two images resemble each other. We used two quantitative methods for validating the registration accuracy: a voxel-based and a landmark-based.

The voxel-based registration accuracy method visualizes how the registration algorithm performs; we only show the most extreme image as it is the most difficult to match. Given that the 6-week-old normal atlas was chosen as reference frame, the most extreme deformation would be registering the atlas to a small 4-week-old Crouzon mouse with the most abnormal skull shape. If the experiment fails, the Crouzon mouse will not resemble the 6-week-old normal atlas with respect to shape.

The registration accuracy was further examined using landmarks. [158] stated that, unlike most evaluation methods, a dense set of landmarks is the ideal way of detecting incorrect registrations. A clinical expert familiar with landmarking, annotated the set of images with 49 anatomical landmarks according to Figure 12.2. The landmarks on the 6-week-old normal atlas were propagated to all subjects using the previously obtained transformations.

Subsequently, landmark errors were estimated. The errors are defined by the point-to-point error, i.e. the Euclidean distance between an automatically obtained landmark and the corresponding manual landmark.

12.2.2.5 Statistical Significance

To enable a statistical analysis of the growth, knowledge of detailed anatomical correspondence had to be established between all mice. The 6-week-old normal atlas was chosen as the common reference frame because of the large variation in the Crouzon data [133]. Subsequently, the 30 specimens were registered to the 6-week-old normal atlas using rigid and non-rigid transformations.

Two different approaches were used for determining whether the estimated growth was significant. The first approach was a local False Discovery Rate (FDR). Local FDR is a method for large-scale simultaneous hypothesis-testing problems, which occur when testing for statistical significance in multiple correlated findings (magnitude of the growth vectors). The second approach was a Multivariate Analysis of Variance (MANOVA). Multivariate Analysis of Variance (MANOVA) is used in cases where there is more than one correlated dependent variable, and where the dependent variables cannot simply be combined. The objective of the MANOVA was to test whether the means for the four mouse groups were taken from the same sampling distribution. This required that we identified whether changes in the independent variables (x, y, and z-directions) had a significant effect on the dependent variables (age, type) and the possible interactions between age and type. A less known problem of the MANOVA is that the sample covariance matrix is not a good estimate of the true population covariance matrix when the number of data points is more than the number of samples [169].

Local FDR
Efron's approach [57] aims at improving the estimation of the false discovery rate (FDR) [16] by replacing the theoretical null-hypothesis by an estimated empirical null-hypothesis. Given (uncorrected) p-values at each point, their corresponding z-values are estimated,

$$z_i = \Phi^{-1}(p_i),\ i = 1, ..., M, \tag{12.3}$$

where Φ is the standard normal cumulative distribution function and M is the number of points under consideration (i.e. all vertices of a surfaces or all voxels in an image). The theoretical null-hypothesis states that z_i $N(0, 1)$. Now, local false discovery rate can be estimated at each point to distinguish "interesting" (significant) values from "uninteresting". This is defined by

$$fdr(z) \equiv \frac{f_0(z)}{f(z)}, \tag{12.4}$$

where $f(z)$ is the value of a curve fitted to the z-histogram (including both classes, interesting and uninteresting) and $f_0(z)$ is the null-hypothesis density.

Instead of using the theoretical null-hypothesis for $f_0(z)$, Efron proposed to use the empirical null-hypothesis, obtained by estimating the mean and standard deviation from $f(z)$. Now, the input values may be classified as interesting or uninteresting by assigning a certain threshold to the fdr-values ([57] suggests that $fdr(z_i) <= 0.10$ gives an interesting finding).

In this study, the growth values (magnitude of the growth vectors) were used as input to the local FDR. Since there is a vast amount of findings (945.284 growth values), it is an important issue to correct for correlations between the findings. A test for significant difference with respect to age was performed on the growth values, such that for each mouse-type (Crouzon or normal) the mean magnitude of the 4-week-old group was compared to that of the 6-week-old group.

MANOVA
Given that the data are mutually independent, identically distributed (same probability function) and Gaussian, a linear model can be set up. It is assumed that every observation (mouse) $\bar{\boldsymbol{X}}_{ijv}$ can be written as

$$\bar{\boldsymbol{X}}_{ijv} = \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{a}}_i + \bar{\boldsymbol{b}}_j + \bar{\boldsymbol{ab}}_{ij} + \bar{\boldsymbol{\epsilon}}_{v(ij)}, \tag{12.5}$$
$$\sum \bar{\boldsymbol{a}}_i = 0, \sum \bar{\boldsymbol{b}}_j = 0, \sum_i \bar{\boldsymbol{ab}}_{ij} = \sum_j \bar{\boldsymbol{ab}}_{ij} = 0,$$

where $\bar{\boldsymbol{a}}_{\boldsymbol{i}}$ denotes the age, $\bar{\boldsymbol{b}}_{\boldsymbol{j}}$ the mouse-type and $\bar{\boldsymbol{ab}}_{\boldsymbol{ij}}$ the interaction between age and mouse-type. Table 12.2 shows the scheme for carrying out the MANOVA. The Mahalanobis distance was used to test for equal means between wild-type and Crouzon mice as well as for equal means between 4-week-old mice and 6-week-old mice. The interaction was tested with a Hotelling's T^2. For more details on the statistics, see the Supplementary Material 12.4.

An advantage of using a MANOVA instead of local FDR was that we could separate which significant areas were related to mouse-type, which were age-related and which were related to both type and age.

Unlike local FDR, all four mouse groups could be tested simultaneously with the MANOVA. This testing determined whether the four group means are equal and determined the spatial location of any significant differences (see e.g. [60]).

12.2.3 Results

The concept of voxel-based accuracy quantification is illustrated in Figure 12.3; here the Crouzon mouse is registered to the atlas using rigid and non-rigid transformations.

The landmarking errors are shown in Figure 12.4 using box and whisker plots.

12.2.3.1 Qualitative Assessment and Visualization of Growth

Within each mouse group, the 6-week-old atlas was registered rigidly to the 4-week-old atlas, using rotations and translations to align the mice. The growth between the 4-week-old and the 6-week-old atlases was constructed by using non-rigid image registration with the 4-week-old atlas as target. Afterwards, the 6-week-old atlas surfaces with point correspondence to the 4-week-old atlas surfaces were extracted using Algorithm 1.

Using linear interpolation, animations showing the craniofacial growth in the wild-type mice and the Crouzon mice were constructed. Lateral views from the animations are shown in Figure 12.5. As 3D animations do not perform well on paper, local deformations were obtained at each point of the surface of the 4-week-old atlases i.e. the deformations from the non-rigid image registration were visualized as color-coded growth vectors, see Figure 12.6.

12.2.3.2 Quantitative Assessment of Average Growth

Figure 12.7 illustrates the significant changes in growth magnitude between 4-week-old mice and 6-week-old mice.

Results from the MANOVA are shown in Figure 12.8.

12.2.3.3 Landmark-based Measurement of Growth

To validate our statistical findings we carried out an evaluation based on 18 skull parameters using the landmarks from Figure 12.2. All individual landmarks were automatically placed using atlas landmarks as explained in Registration Accuracy 12.2.2.4, Table 12.3.

Left and right sides of the mid-sagittal plane were assumed to be similar and the average of each skull parameter was calculated. Table 12.4 and Figure 12.9 show the average skull parameters for the four mouse groups measured in millimeters along with the calculated growth rates (slopes). As a consequence of the unbalanced cross-sectional data, it was not possible to generate a one-to-one correspondence for estimation of the growth rates. Hence, the growth rates are based on averaging all combinations of 4-week-old and 6-week-old mice within

each group. The growth rates were tested for equal means using a t-test ($\alpha = 5\%$).

12.2.4 Discussion

The growth model is affected by the choice of coordinate system used when specifying the growth vectors v. While the non-rigid image registration assures that the vectors represent actual displacements of corresponding anatomical points, the relative alignment of the structures in a pair of images affects the length and orientation of v. Coordinate systems may, in the context of image registration, be divided into two types: anatomical and mathematical.

A popular anatomical coordinate system used when comparing human skulls is the reference system based on the mid-sagittal (vertical) plane (MSP) and the Frankfort horizontal plane (FHP). This system aligns a pair of skulls such that the lines between the ears coincide, and the planes containing the ears and the lower orbital rim coincide. As a consequence, a measured difference between the skulls (e.g. due to growth), will tend to be small at locations close to the MSP and FHP (especially at the midpoint between the ears) and larger at locations farther away. A growth model built using this system will be biased, but will have the advantage of potentially being easier to interpret by clinicians and researchers accustomed to this particular coordinate system. Transforming the displacement vectors between different anatomical coordinate systems is straightforward.

An example of a mathematical coordinate system is one that is defined by giving all parts of the image equal weights (e.g. in a least-squares sense) during alignment. So-called Procrustes alignment is of this kind and is often used in analysis of shape [23, 53] and in image analysis systems for shape recognition [42]. In principle, it is possible to transform displacement vectors from a mathematical coordinate system to an anatomical coordinate system, and vice versa.

A drawback of any particular choice of coordinate system is that its use may lead to different results and interpretations than if a different coordinate system had been chosen. Alternative methods, not based on direct analysis of spatial coordinates of points, but on distances between points, are invariant to the coordinate system and, thus, avoid this drawback. As an example, the position of the eye varies as the head rotates, but the distance between the eyes does not. Methods in X-ray cephalometry often define variables based on distance between pairs of landmarks [181, 80] and growth may be defined in terms of the ratio between such distances measured at subsequent times. A limited number of variables are typically defined in relation to the biological or clinical

problem under investigation. A systematic, coordinate-system invariant approach to comparison of shapes based on distances is the Euclidean distance matrix analysis (EDMA) method by Lele and Richtmeier [109], where all possible pairs of distances between a sparse set of landmarks are analyzed.

In the present work, a mathematical coordinate system was chosen. It is inherent in the image registration process and is determined by the rigid registration, using the whole image, of each mouse to the reference atlas before the non-rigid part of the registration is carried out. The growth model was built using this particular coordinate system, but provided also the possibility to analyze the growth of distances between any pair of anatomical points in a coordinate system invariant fashion. In fact, the establishment of detailed point correspondences across the population of mice, as determined by the image registration, provides a unique possibility to carry out a quasi-EDMA on a dense set of point distances.

12.2.4.1 Registration Accuracy

Judged from Figure 12.2, the registration of the most extreme 4-week-old Crouzon mouse to the normal 6-week-old atlas was successful. There were minor differences between the atlas and the transformed Crouzon mouse in the nasal region and posterior part of the skull. The slightly discoloring around the bones in the transformed Crouzon mouse is a result of interpolation. According to [158] landmark annotation is a much more robust measure for detecting misregistrations than visualizations, image similarity, differences images, etc.

The registration accuracy from the growth modeling was measured by a comparison of manually placed landmarks and automatically obtained landmarks. Fewer outliers were present in the dataset for 4-week-old mice, maybe because of the smaller dataset. With most registration errors below 0.5 mm, the accuracy was within the range of previously estimated errors on a similar dataset [133]. Testing the automatic assessments versus the manual assessments in a t-test (5% level of significance) revealed equal means for the non-rigid and manual approach.

12.2.4.2 Visualizations of Atlas-based Growth

Clinical experts evaluated the growth model (Figures 12.5 and 12.6) according to existing knowledge. The animation of Crouzon mouse growth showed growth inhibition in the nose and front of calvaria (probably related to the coronal sutures) and in the posterior part of the calvaria (probably related to the

lambdoid sutures or possibly to sutures and synchondroses in the cranial base). Also, the skull width expanded. These findings support research by [102]. The orbital rims expanded outwards for both types. Watching the animations in a lateral view, the animation of Crouzon mouse growth expanded in height in the calvaria and the nasal region. The nose in the animation of normal mouse growth elevated and extended forwards because of the growth in the sutures. In the animation of Crouzon mouse growth, the nose did not elevate at all and extended by a very small amount. Although the nose in the animation of Crouzon mouse growth did not extend much, and did not elevate, the zygoma extended forward by a large amount; it also elevated slightly at the anterior end, creating a pronounced bulge between the eyes. The bulge was not present at 4 weeks, but became very pronounced at 6 weeks. This could be related to adaptation due to muscle function, i.e. apposition of bone in the vicinity of muscle attachment. The mandible did not move as much forward in the animation of Crouzon mouse growth as in the animation of normal mouse growth. The posterior edge of the mandible stayed in the same place relative to the ear/posterior part of calvaria, while in the normal mice this distance increased. This finding is probably related to fusion of the spheno-occipital synchondrosis in some of the Crouzon mice. All deviations in growth are interpreted as dysplastic and compensatory growth caused by the fusion of sutures and synchondroses.

12.2.4.3 Local FDR

Both the atlas-based growth and local FDR compare changes in each mouse group as a function of age, which makes visual comparison between the method outcomes possible. However, local FDR uses all specimens for finding significance, whereas the atlas-based approach is measured on group averages. The significant normal growth areas in Figure 12.7 agree with what we would expect from looking at the visualization of average growth in Figure 12.6. The significant findings at the incisors might be related to variations in tooth position secondary to variations in maxillary development. In the Crouzon case, the local FDR analysis does not find the growth in the posterior part of the head, mid-calvaria and at the transition between nasal and frontal bone significant. However, the growth of the orbital rims, the ridges on the frontal bone and the mandibular growth is rated significant. The study by [133] showed that the group of Crouzon mice had a larger inter-subject variance compared to the group of normal mice. Thus, an explanation of the discrepancy regarding the local FDR result on the significant Crouzon growth could be related to the size of the dataset. As the normal mice had less in-group variation, their growth pattern was found to be more consistent, whereas the dataset size was not large enough to deal with the large variation in Crouzon mice.

12.2.4.4 MANOVA

A feature of the MANOVA design was that we could find the significant areas where growth differed for mouse type, age and the interaction between age and mouse type. When testing for significant areas regarding differences as a result of mouse type, the Crouzon mice were tested against normal mice regardless of age. Similarly, Crouzon mice and normal mice were merged according to age when testing for differences between 4 and 6 weeks. The significant areas regarding mouse type confirm our expectations of a wider, higher, shorter skull in the Crouzon mice compared to the normal mice. The significant areas at the incisors are probably related to the reduced forward growth of the maxilla in the Crouzon mice compared to the normal mice. The age results were more subtle, the shape of the zygomatic bone was different from Crouzon to normal mice when looking at the data. However, there was a size difference even between 4-week-old and 6-week-old Crouzon mice. The only interpretable interaction result was the location of the dots, Figure 12.8, where the lower orbital rim connects to the nose. This corresponds to the bulge which evolved in the Crouzon mice between week 4 and 6, as seen in the animations.

12.2.4.5 Skull Parameters

Automation of the annotation process helped reducing errors and limit manual work. However, landmark-based measures will still provide sparse information. With the statistical methods we were able to pinpoint which landmarks would be interesting to use for measuring skull parameters; while the skull parameters provided measurable details about the morphology. The dents on the nose would, however, be hard to measure using landmarks.

All results in the following paragraph were tested as statistically significant group-wise using the Wilcoxon rank sum test (5% level of significance), see 12.4 and Figure 12.9. Skull length (SL), nose length (NL), mandible length to 1st molars (MLM), mandible length 1-3 (MW1-3), and zygomatic arch (Z) were significantly decreased for Crouzon mice. Nose width (NW) was significantly increased for Crouzon mice. For skull height (SH), mandible width 1 and 3 (MW1, MW3), mandible height (MH), interorbital distance ([140]; IODp) and inner canthal distance (ICD), the measures on the 4-week-old mice were found similar regardless of mouse-type. However, the development to week 6 was different: MH and ICD were found to have equal growth rate, since the 4-week-old mice were rated equal and similar for the 6-week-old mice; SH, MW1, and IODp increased for Crouzon mice, whereas MW3 decreased for Crouzon mice. No significant changes were found for maxilla length (MX) and interorbital distance

(IOD) in normal mice; in Crouzon mice MX decreased and was receding while IOD increased. Skull width (SW) was tested to be equal for 6-week-old normal mice and 6-week-old Crouzon mice. Mandible width 2 (MW2) was tested as equal for 4 and 6-week-old normal mice; and equal for 6-week-old normal mice and 4-week-old Crouzon mice. However, the 4-week-old group could not be considered equal. Looking at the growth rates for SL and NL, the Crouzon mice grew at approximately half the normal rate. SH, SW, and IOD are approximately twice the normal rate. MX growth rate was approximately eight times larger for normal mice; the negative rate for the Crouzon mice is probably a result of reduced maxillary sutural growth. MW2 growth rate was almost three times larger for Crouzon mice. MW1, MH, IODp were almost a quarter larger than the normal rate. Crouzon mice were inhibited in growth for MLM, ML1, and Z, while NW growth increased compared to normal growth. Finally, MW3, ML2, ML3 and ICD had equal growth rates according to the Student's t-test.

As expected, the Crouzon mice displayed inhibited growth in e.g. SL, NL and MX and increased growth in e.g. SH, SW, NW. These findings are in agreement with observations on craniofacial growth in humans with Crouzon syndrome [102].

12.3 Concluding Remarks

We have built a growth model for automatic phenotyping using appropriate registration models. This growth model confirmed pre-existing knowledge of Crouzon development in humans to be similar in mice.

The Crouzon mice were clearly different from a control group of wild type mice in terms of growth, a difference caused by the fused sutures; e.g. skull and nose length were inhibited. Skull height, for instance, was rated equal for 4-week-old mice; while at 6 weeks, the skull height had increased for Crouzon mice. Other measures, such as skull width, revealed compensatory growth in the Crouzon group. With the automatically placed landmarks, we were able to estimate the growth rates for both Crouzon and normal mice. For example, skull length and nose width grew almost half as much for Crouzon mice compared to normal mice. On the other hand, skull width and interorbital distance grew twice as much for Crouzon mice.

Future work could include a wider range of mouse ages as well as an assessment of growth in a longitudinal study on human data. Since the method is automatic, landmarks are easier to obtain and we are able to measure every desired parameter, irrespective of its location on the mouse.

12.3.1 Acknowledgments

The authors would like to thank Professor Gillian Morriss-Kay, University of Oxford, UK, for supervision and valuable advice in relation to breeding of the mice. Also thanks to PhD, M.Sc. (Eng.) Karl Sjöstrand for his Matlab-implementation of local FDR.

For all image registrations, the Image Registration Toolkit was used under License from Ixico Ltd.

For surface visualizations and landmarking etc., the VTK-based tool 'landmarker' was used. [48].

12.4 Supplementary Material

12.4.1 Multivariate Analysis of Variance

Multivariate Analysis of Variance (MANOVA) is used in cases where there are more than one correlated dependent variable and where the dependent variables cannot simply be combined. The objective is to identify whether changes in the independent variables have a significant effect on the dependent variables as well as identify the possible interactions among the independent variables and the connection among dependent variables. By splitting the dataset into four groups depending on age and mouse-type, it is possible to test whether the group means are different and determine spatial location of any significant differences. This has been done according to the MANOVA; see e.g. [60]. Throughout this section vectors and matrices are written in boldface.

An observation is characterized by $\bar{\boldsymbol{X}}_{ijv} = \begin{bmatrix} \mathbf{x}_{ijv} \\ \mathbf{y}_{ijv} \\ \mathbf{z}_{ijv} \end{bmatrix}$ with $i = \begin{cases} 1\text{: normal} \\ 2\text{: Crouzon} \end{cases}$

$j = \begin{cases} 1\text{: 4-week} \\ 2\text{: 6-week} \end{cases}$ and $v = \{1, 2, ..., n_{ij}\}$ or repetitions (number of mice in each group, which is five or ten). $\bar{\boldsymbol{X}}_{ij.}$ denotes the group mean, i.e. $\bar{\boldsymbol{X}}_{11.}$ is the mean of the deformation vectors from the five 4-week normal mice, see an overview of the MANOVA in Table 12.5.

The objective is to determine whether $\bar{\boldsymbol{X}}_{11.}$, $\bar{\boldsymbol{X}}_{12.}$, $\bar{\boldsymbol{X}}_{21.}$, $\bar{\boldsymbol{X}}_{22.}$ are dissimilar from each other. $\boldsymbol{X}$ denotes a matrix consisting of all $\bar{\boldsymbol{X}}_{ij.}$'s, in this case a $30 \times 3 \times 945.284$ matrix (30 mice, 3 growth directions x, y, z and 945.284 vertices).

Assuming that the data is mutually independent, identically distributed (same probability function), and Gaussian, a model can be set up. It is assumed that every observation $\bar{\boldsymbol{X}}_{ijv}$ is normal distributed with same variance-covariance matrix $\boldsymbol{W}$ and mean values

$$\bar{\boldsymbol{X}}_{ijv} = \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{a}}_i + \bar{\boldsymbol{b}}_j + \bar{\boldsymbol{ab}}_{ij} + \bar{\boldsymbol{\epsilon}}_{v(ij)}, \tag{12.6}$$

where the parameters $\bar{a}_i$, $\bar{b}_j$ and $\bar{ab}_{ij}$ satisfy

$$\sum \bar{\boldsymbol{a}}_i = 0, \ \sum \bar{\boldsymbol{b}}_j = 0, \ \sum_i \bar{\boldsymbol{ab}}_{ij} = \sum_j \bar{\boldsymbol{ab}}_{ij} = 0$$

where $\boldsymbol{b}$ denotes the age, $\boldsymbol{a}$ the mouse-type and $\boldsymbol{ab}$ the interaction between age and mouse-type. A desirable model property would be if the interaction term could be discarded as growth would then only be represented by variable $\boldsymbol{b}$, hence, a test is required.

The relationship is described in Table 12.6.

The expected values of $\bar{\boldsymbol{X}}_{11.}$, $\bar{\boldsymbol{X}}_{12.}$, $\bar{\boldsymbol{X}}_{21.}$, $\bar{\boldsymbol{X}}_{22.}$ can be estimated from the model, Equation 12.6. The model in matrix form

$$\begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{ab}} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{X}}_{\mathbf{11.}} \\ \bar{\boldsymbol{X}}_{\mathbf{12.}} \\ \bar{\boldsymbol{X}}_{\mathbf{21.}} \\ \bar{\boldsymbol{X}}_{\mathbf{22.}} \end{bmatrix} \tag{12.7}$$

A beneficial property of the 2 × 2-design is that it is fairly simple, because all tests ($\bar{\boldsymbol{a}} = \bar{\mathbf{0}}$, $\bar{\boldsymbol{b}} = \bar{\mathbf{0}}$, $\bar{\boldsymbol{ab}} = \bar{\mathbf{0}}$) only consider the variation in one dimension, e.g. to test for $\bar{\boldsymbol{a}} = \bar{\mathbf{0}}$ the difference is measured between ($\bar{\boldsymbol{X}}_{11.} + \bar{\boldsymbol{X}}_{12.}$) and ($\bar{\boldsymbol{X}}_{21.} + \bar{\boldsymbol{X}}_{22.}$). Testing if $\bar{\boldsymbol{b}} = \bar{\mathbf{0}}$ (age) the difference is measured between ($\bar{\boldsymbol{X}}_{11.} + \bar{\boldsymbol{X}}_{21.}$) and ($\bar{\boldsymbol{X}}_{12.} + \bar{\boldsymbol{X}}_{22.}$), while testing $\bar{\boldsymbol{ab}} = \bar{\mathbf{0}}$ the difference is measured between ($\bar{\boldsymbol{X}}_{11.} + \bar{\boldsymbol{X}}_{22.}$) and ($\bar{\boldsymbol{X}}_{12.} + \bar{\boldsymbol{X}}_{21.}$). The difference is measured with respect to the residual variance-covariance matrix $\boldsymbol{W} = \frac{1}{m}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T$ using the Mahalanobis Distance,

$$D^2 = (\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2)\boldsymbol{W}^{-1}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2)^T. \tag{12.8}$$

$\boldsymbol{\mu}$ was found from the model in Equation 12.7 and $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ denotes $\bar{\boldsymbol{X}}_{1..}$ and $\bar{\boldsymbol{X}}_{2..}$ when comparing type and $\bar{\boldsymbol{X}}_{.1.}$ and $\bar{\boldsymbol{X}}_{.2.}$ when comparing age. The interaction term cannot be estimated with the Mahalanobis Distance, but an explanation on how the interaction term was estimated will follow.

Assuming the same number of mice n in each group m, the degrees of freedom, are $4n-4$, where $4n-4$ refers to the four groups with an assumed equal number of mice n. The -4 refers to the removal of the mean within each of the four groups. The test used with the Mahalanobis distance is equivalent to Hotelling's T^2.

The null-hypothesis: The interaction between mouse-type $\boldsymbol{a}$ *and age* $\boldsymbol{b}$ *is zero, or* $H_0 : \hat{\boldsymbol{ab}} = \bar{\mathbf{0}}$.
Using Hotelling's T^2 the significance of the interaction can be found:

$$t^2 = 4n\, \hat{\boldsymbol{ab}}^T \boldsymbol{W}^{-1} \hat{\boldsymbol{ab}} \in T^2(p, m), \tag{12.9}$$

with $p = 3$ each deformation vector has an x, y and z coordinate. The T^2-distribution can be converted to a F-distribution using the conversion term $\frac{m-p+1}{pm} t^2 \in F(p, m-p+1)$. Applying the conversion factor to the T^2-test Equation 12.9 the corresponding F-test is found

$$\begin{aligned} \frac{4n-4-3+1}{3(4n-4)} 4n\, \hat{\boldsymbol{ab}}^T \boldsymbol{W}^{-1} \hat{\boldsymbol{ab}} &\in F(p, m-p+1) \\ \frac{4n-6}{12n-12} 4n\, \hat{\boldsymbol{ab}}^T \boldsymbol{W}^{-1} \hat{\boldsymbol{ab}} &\in F(p, m-p+1) \end{aligned} \tag{12.10}$$

However, none of the two tests are ready for use in this study because the material is unbalanced - the 4-week-old material has 5 mice in each group and the 6-week-old material has 10 mice in each group. Hence, to test the interaction term in this unbalanced design the degrees of freedom have to be adjusted. The degrees of freedom m in the unbalanced case is $m = n_{11} + n_{12} + n_{21} + n_{22} - 4$ i.e. the number of mice with subtraction of the mean. The testing done with an F-test where the number of mice and the conversion term from a T^2 to a F-test has been corrected due to the change in m and n.

$$\begin{aligned} \frac{\sum n_{ij} - 6}{3\sum n_{ij} - 12} \sum n_{ij}\, \hat{\boldsymbol{ab}}^T \boldsymbol{W}^{-1} \hat{\boldsymbol{ab}} &\in F(p, \sum n_{ij} - 6) \\ \frac{30-6}{90-12} \sum n_{ij}\, \hat{\boldsymbol{ab}}^T \boldsymbol{W}^{-1} \hat{\boldsymbol{ab}} &\in F(p, \sum n_{ij} - 6) \\ \frac{24}{78} \sum n_{ij}\, \hat{\boldsymbol{ab}}^T \boldsymbol{W}^{-1} \hat{\boldsymbol{ab}} &\in F(p, \sum n_{ij} - 6) \end{aligned} \tag{12.11}$$

Choosing a suitable level, e.g. $\alpha = 5\%$, points are significant when $1 - \alpha > 0.95$.

12.4.2 Landmark Definitions for Skull Parameters

1. Most anterior point on the nasal bone.
2. Most anterior, inferior midpoint between the central incisors.
3. Most posterior midpoint on fronto-nasal suture.
4. Midpoint of premaxillary suture on the right side.
5. Midpoint of premaxillary suture on the left side.
6. Most medial incisal point on the right mandibular incisor.
7. Most medial incisal point on the left mandibular incisor.
8. Most posterior point on the left mandibular angle.
9. Most posterior point on the right mandibular angle.
10. Upper most posterior point on the right side of the mandible.
11. Upper most posterior point on the left side of the mandible.
12. Upper most point on the left coronoid process.
13. Upper most point on the right coronoid process.
14. Intersection of parietal, interparietal, and occipital suture on the left side.

15. Intersection of parietal, interparietal, and occipital suture on the right side.
16. Most right posterior point on the interparietal-occipital suture.
17. Most left posterior point on the interparietal-occipital suture.
18. Most left point on the external foramen magnum.
19. Most right posterior point on the interparietal-occipital suture.
20. Most right point on the external foramen magnum.
21. Right periodic capsule.
22. Left periodic capsule.
23. Most anterior point on the external foramen magnum.
24. Midpoint of occipital and basis-sphenoid suture.
25. Most posterior point of the right internal pterygoid process.
26. Most posterior point of the left internal pterygoid process.
27. Point on the right mandibular process next to 1. molar.
28. Point on the left mandibular process next to 1. molar.
29. Midpoint on palatine process.
30. Right orbital foramen.
31. Left orbital foramen.
32. Point on the skull where the frontal and parietal suture intersects.
33. Point on the skull where the parietal and interparietal suture intersects.
34. Upper most part on the external part of foramen magnum.
35. Point on the left tympanic bone.
36. Point on the right tympanic bone.
37. Point on the left maxillary process next to the 1. molar.
38. Point on the right maxillary process next to the 1. molar.
39. Most anterior point on the mandible.
40. Most anterior point on the nasal spine.
41. Right premaxilla.
42. Left premaxilla.

43. Left maxillary tuber.
44. Right maxillary tuber.
45. Most anterior point on the presphenoid bone.
46. Most anterior point on the left zygomatic arch.
47. Most posterior point on the left zygomatic arch.
48. Most anterior point on the right zygomatic arch.
49. Most posterior point on the right zygomatic arch.

12.5 Tables, Figure Text, Figures

Algorithm 1 - Detailed point correspondences for growth	
1:	Register all mice (sources) to the chosen atlas (target) using rigid and non-rigid transformations to obtain a voxel-wise non-rigid transformation for each source.
2:	**repeat**
3:	Extract a skull surface from the target.
4:	Use the non-rigid transformation on every target surface vertex to warp the vertex from the target surface to the new surface resembling the source mouse.
5:	**until** point correspondences have been built for all mice (sources).

Table 12.1: Schematic method.

E{ }	B_1 (4-week)	B_2 (6-week)
A_1 (Wild-type)	$\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{a}} - \bar{\boldsymbol{b}} - \bar{\boldsymbol{ab}}$	$\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{a}} + \bar{\boldsymbol{b}} + \bar{\boldsymbol{ab}}$
A_2 (Crouzon)	$\bar{\boldsymbol{\mu}} + \bar{\boldsymbol{a}} - \bar{\boldsymbol{b}} + \bar{\boldsymbol{ab}}$	$\bar{\boldsymbol{\mu}} + \bar{\boldsymbol{a}} + \bar{\boldsymbol{b}} - \bar{\boldsymbol{ab}}$

Table 12.2: Expected mean values found, based on the model in equation 12.6. The model assumes that every mouse can be constructed as an overall mean plus/minus age, mouse-type and interaction between age and type.

Skull parameter abbreviations	
SL	Skull length
NL	Nasal length
SH	Skull height
MX	Maxilla length (left/right)
MW1-3	Mandible width 1-3
MH	Mandible height (left/right)
MLM	Mandible length to 1st molars (left/right)
ML1-3	Mandible length 1-3 (left/right)
SW	Skull width
NW	Nose width
IOD	Interorbital distance
IODp	Interorbital distance ([140])
ICD	Inner canthal distance
Z	Zygomatic arches (left/right)

Table 12.3: Abbreviated skull parameters.

∗ indicate combined left and right		Landmarks	Mouse type				Growth rate		
			4-week Normal	6-week Normal	4-week Crouzon	6-week Crouzon	Normal	Crouzon	Difference %
Skull parameters [mm]	SL	34 - 40	20.56	23.60	18.16	19.79	3.04	1.66	45.4
	NL	40 - 45	13.24	15.49	11.20	12.46	2.25	1.26	44.1
	SH	16 - 24	†7.21	7.59	†7.22	7.99	0.38	0.81	114.3
	MX*	37/38 - 43/44	†3.78	†3.80	3.47	3.31	0.02	-0.15	789.7
	MW1	8 - 9	†8.29	9.39	†8.31	9.69	1.11	1.39	25.9
	MW2	10 - 11	‡10.09	††10.27	†10.38	11.06	0.18	0.67	275.5
	MW3	12 - 13	†9.00	9.60	†8.76	9.39	0.60	0.53	♯ 11.6
	MH*	8/9 - 11/10	†3.36	‡3.92	†3.25	‡3.93	0.57	0.70	23.7
	MLM*	11/10 - 28/27	7.64	8.66	7.11	7.93	1.02	0.86	15.2
	ML1*	8/9 - 39	10.24	11.93	9.47	10.83	1.69	1.38	18.5
	ML2*	11/10 - 39	10.91	12.48	10.30	11.77	1.56	1.51	♯ 3.4
	ML3*	12/13 - 39	8.67	10.03	7.99	9.34	1.35	1.31	♯ 3.2
	SW	35 - 36	9.86	†10.14	†10.16	10.72	0.28	0.58	106.0
	NW	41 - 42	3.03	3.49	3.23	3.87	0.47	0.73	56.3
	IOD	30 - 31	†5.62	†5.82	6.13	6.58	0.20	0.47	131.7
	IODp	46 - 48	†6.03	6.64	†6.03	6.80	0.61	0.77	27.8
	ICD	4 - 5	†4.17	‡4.50	†4.20	‡4.50	0.33	0.32	♯ 4.8
	Z*	46/48 - 47/49	7.04	8.05	6.01	6.66	1.01	0.62	37.9

Table 12.4: Average skull parameters measured from manually and automatically placed landmarks. The landmarks column shows which landmarks were used in the distance calculation. Row-wise identical symbols (‡,†) indicate equal medians for the Wilcoxon rank sum test ($\alpha = 0.05$). The growth rates are calculated from averaging all combinations of 4-week-old and 6-week-old mice within each group. Growth rates marked by ♯ were tested equal using a Student's t-test ($\alpha = 5\%$). Percentages are calculated with the normal rate as the base.

	B_1 (4-week)	B_2 (6-week)	
A_1 (Wild-type mice)	n_{11} $\bar{X}_{11.}$	n_{12} $\bar{X}_{12.}$	$n_{11} + n_{12}$ $\bar{X}_{1..}$
A_2 (Crouzon mice)	n_{21} $\bar{X}_{21.}$	n_{22} $\bar{X}_{22.}$	$n_{21} + n_{22}$ $\bar{X}_{2..}$
	$n_{11} + n_{21}$ $\bar{X}_{.1.}$	$n_{12} + n_{22}$ $\bar{X}_{.2.}$	$n_{11} + n_{12} + n_{21} + n_{22}$ $\bar{X}_{...}$

Table 12.5: Table for MANOVA testing. $\bar{X}_{11.}$ is the mean of the 4 week normal mice with n_{11} number of mice, similar for $\bar{X}_{12.}$, $\bar{X}_{21.}$, and $\bar{X}_{22.}$. $\bar{X}_{1..}$ is the sum of the two means in the A_1 row, $\bar{X}_{.1.}$ is the sum of the two means in the B_1 column, similar for $\bar{X}_{2..}$ and $\bar{X}_{.2.}$. $\bar{X}_{...}$ is the overall mean e.g. the mean of the means $\bar{X}_{11.}$, $\bar{X}_{12.}$, $\bar{X}_{21.}$ and $\bar{X}_{22.}$. The n's are the number of mice in each group with $n_{11} = n_{21} = 5$ and $n_{12} = n_{22} = 10$ in this study.

E{ }	B_1 (4-week)	B_2 (6-week)
A_1 (Wild-type)	μ - a - b - ab	μ - a + b + ab
A_2 (Crouzon)	μ + a - b + ab	μ + a + b - ab

Table 12.6: Expected mean values for each mouse-group. The idea is that every mouse can be constructed as an overall mean plus/minus age, mouse-type and interaction between age and type.

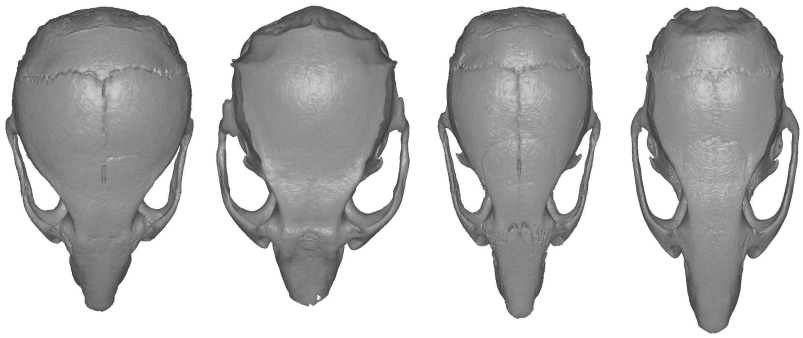


Figure 12.1: Four example mice (from left to right): 4-week-old Crouzon mouse; 6-week-old Crouzon mouse; 4-week-old normal mouse; 6-week-old normal mouse.

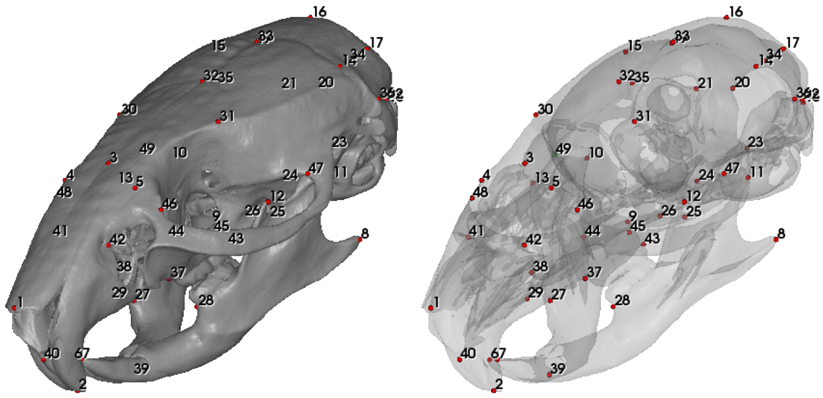
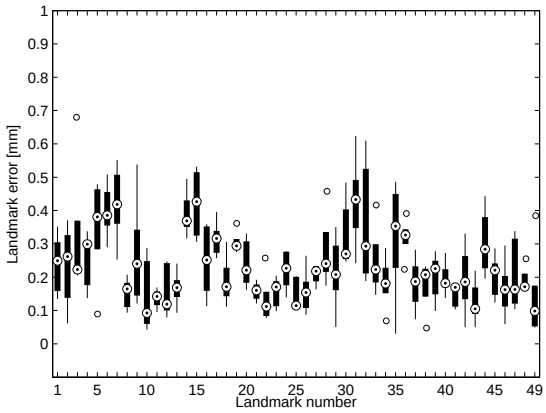
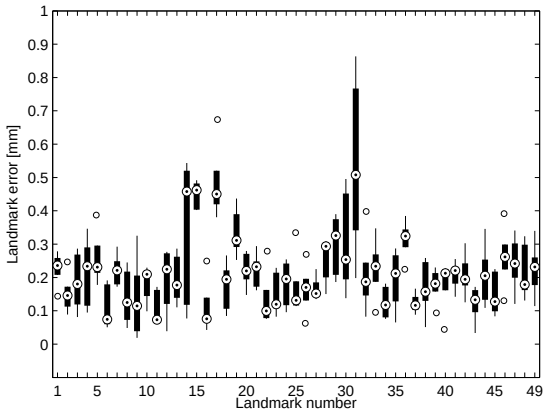


Figure 12.2: The 49 landmarks used for validating the registration - here shown on the 6-week-old normal atlas. Landmark definitions are found in the Supplementary Material 12.4.2.

Figure 12.3: A slice through the micro-CT scans of an extreme 4-week-old Crouzon (left), the 6-week-old wild-type atlas (middle), and the 4-week-old Crouzon deformed to resemble the atlas (right). The images have been inverted to enhance the skull bones.

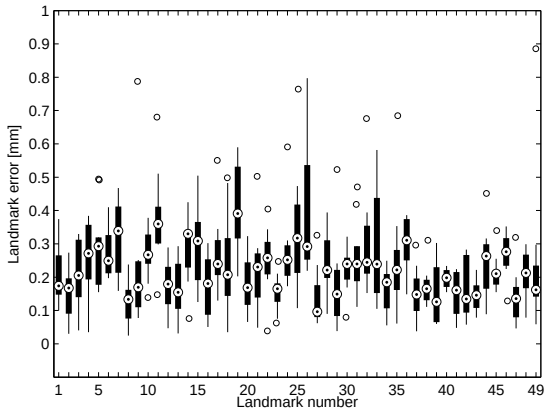


(a) 4-week old Crouzon

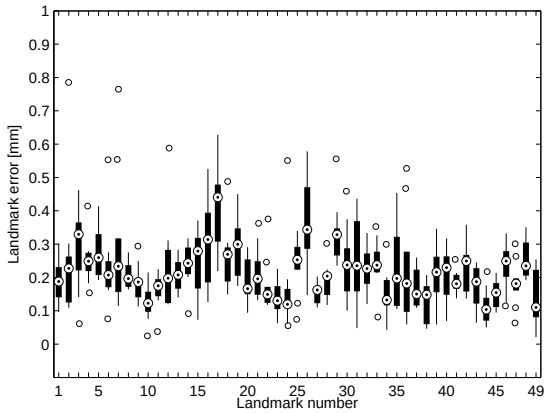


(b) 4-week old normal

Figure 12.4: Landmark errors in millimeter. Manually placed landmarks are compared to automatically obtained landmarks. (a) 4-week-old Crouzon, (b) 4-week-old normal, (c) 6-week-old Crouzon, and (d) 6-week-old normal. The following definition of a box and whisker plot is used here: The box surrounds measurements between the upper and the lower quartile of the data. The symbol $\odot$ inside the box denotes the median of the data. The maximum length of the whiskers is 1.5 times the interquartile range (IQR). Outliers (those lying outside the limits of the whiskers) are marked by a circle.



(c) 6-week old Crouzon



(d) 6-week old normal

Figure 12.4: Continued. Landmark errors in millimeter. Manually placed landmarks are compared to automatically obtained landmarks. (a) 4-week-old Crouzon, (b) 4-week-old normal, (c) 6-week-old Crouzon, and (d) 6-week-old normal. The following definition of a box and whisker plot is used here: The box surrounds measurements between the upper and the lower quartile of the data. The symbol $\odot$ inside the box denotes the median of the data. The maximum length of the whiskers is 1.5 times the interquartile range (IQR). Outliers (those lying outside the limits of the whiskers) are marked by a circle.

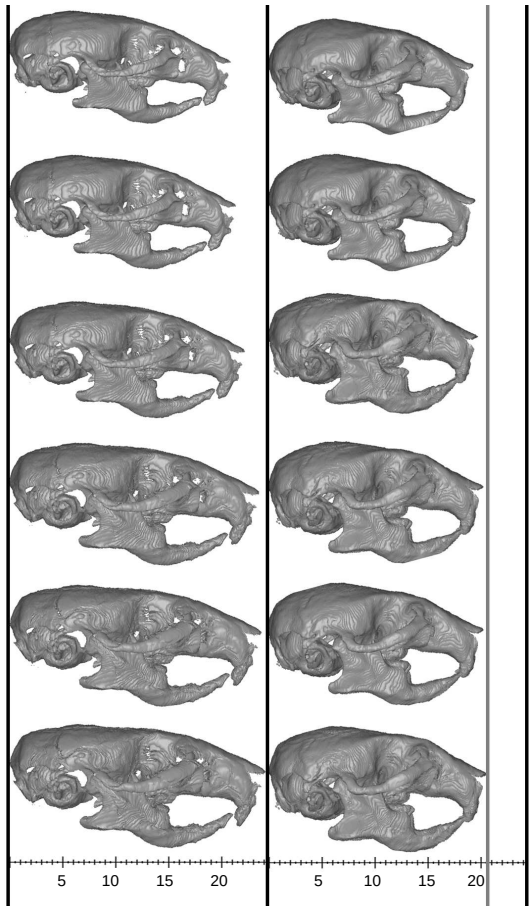


Figure 12.5: Visualization of the normal growth series (left) and the Crouzon growth series (right). The 4-week-old atlas (top) becomes the 6-week-old atlas (bottom) using linear interpolation. Equidistant lines (black) separate the two mouse groups. 6-week-old normal atlas is 24.1 mm long (most posterior point to most anterior point), while the 6-week-old Crouzon atlas measures 20.4 mm.

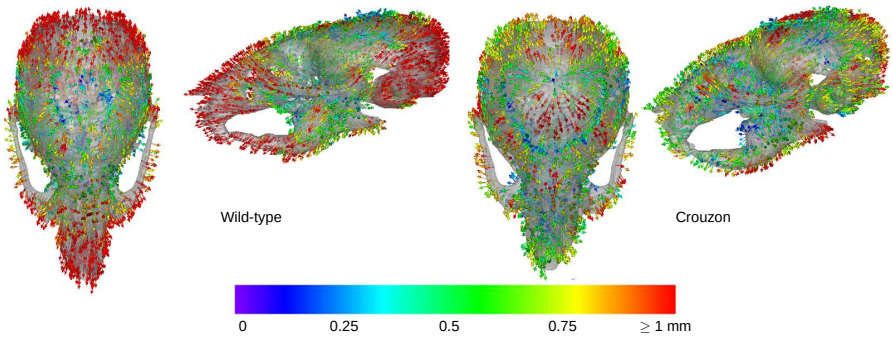


Figure 12.6: Every 200th deformation vector shown on 4-week-old atlases. Left: The 4-week-old normal atlas with growth vectors. Right: The 4-week-old Crouzon atlas. The colors denote displacement with respect to atlas (in mm), with red denoting most change (1 mm and more) and purple no change.

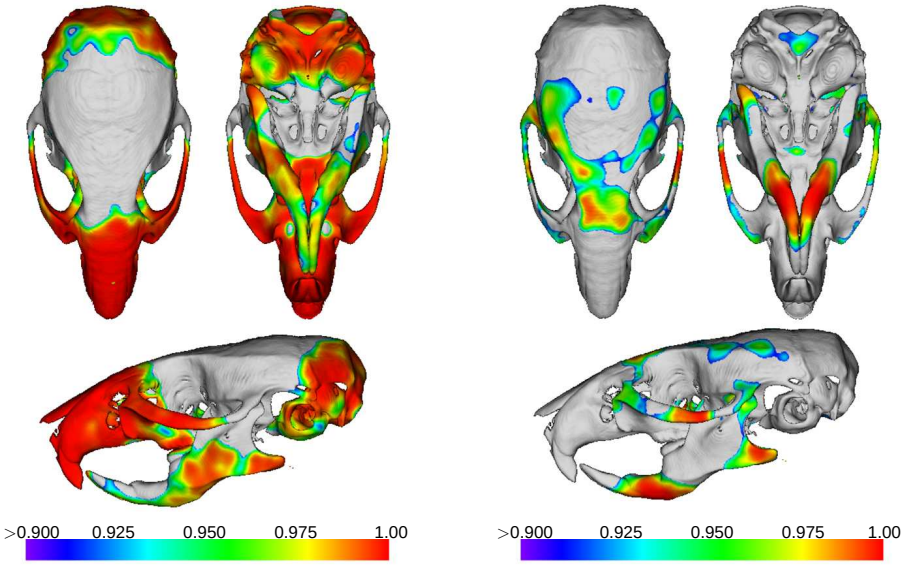


Figure 12.7: The significant areas estimated by local FDR are visualized on the 6-week-old normal atlas to ease the comparison between normal and Crouzon mice. The left image shows that the normal mice change significantly at the nasal region, maxillary bone, premaxilla, occipital, interparietal bone, external auditory meatus, tympanic bulla, malar process, teeth, angle, lower mandible and zygomatic arches, between week 4 and week 6 . The right image shows the areas with significant change between 4-week-old and 6-week-old Crouzon mice. Changes occur in the frontal region, parts of the parietal bone, maxilla, zygomatic bones, angle, articular surface, condyloid process, mandibular foramen, and on the mandible close to the incisors. [41]. 1-fdr displayed on the atlas surface: areas less than or equal to 90% are non-significant (gray), whereas most significant areas are red. Non-significant areas are gray (1-fdr ≤ 0.9) and significant areas (1-fdr > 0.9) are colored from blue (least significant) to red (very significant).

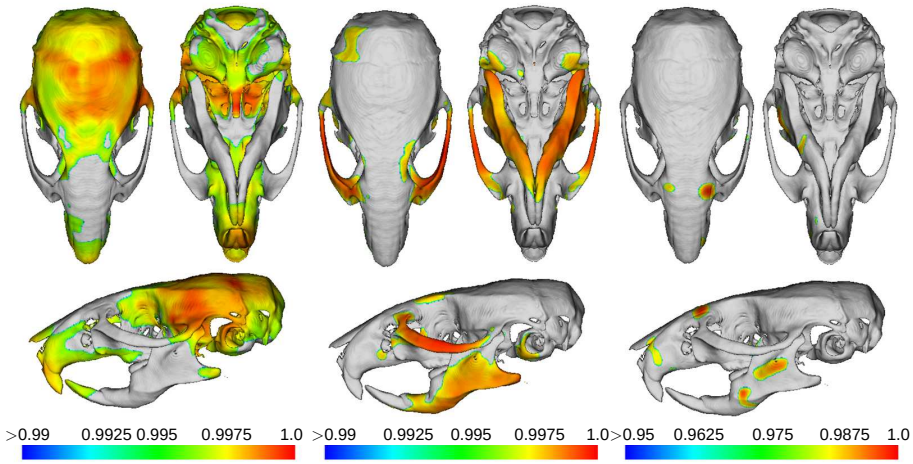


Figure 12.8: MANOVA results visualized on the 6-week-old normal atlas. The left image shows areas with significant difference between normal mice and Crouzon mice regardless of age. The nasal region, incisors, tympanic bulla, squamosal, angle, articular surface, condyloid process, mandibular foramen and the calvaria are in the significant group. The middle image shows the significant differences between 4-week-old and 6-week-old mice regardless of mouse type. The differences are found at the maxilla, zygomatic bone, various places on the calvaria, lower mandible and angle. The right image shows that the interaction between age and type was only significant in a few areas: At the lacrymal, various places on the mandible and in the nasal region. Non-significant areas are gray and significant areas are rated from blue (least significant) to red (very significant). Mouse type and age were tested at a 99% level (alpha = 0.01) and interaction at a 95% level (alpha = 0.5) using an F-test.

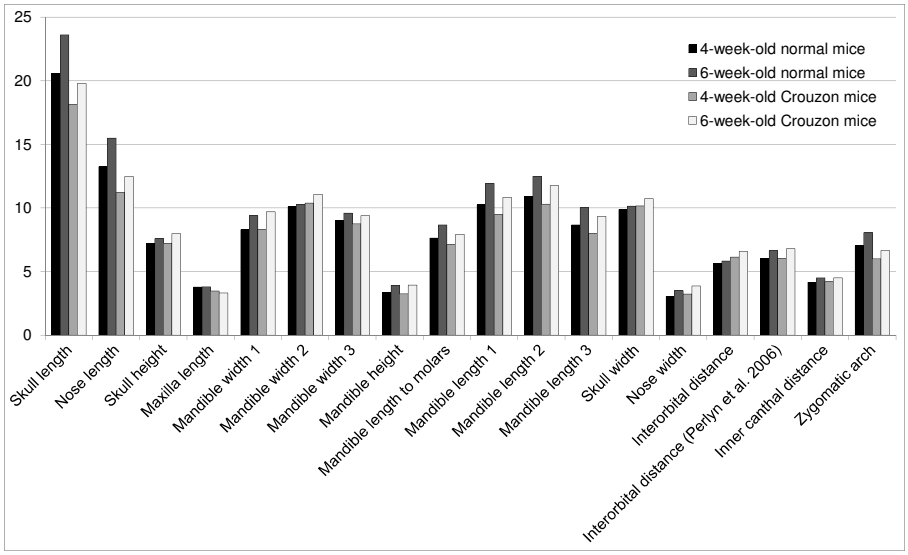


Figure 12.9: Diagram showing the average magnitude of the 18 skull parameters for each mouse-type. All parameters are measured in millimeters [mm]. Black bins denote 4-week-old normal mice, dark gray bins denote 6-week-old normal mice, light gray bins denote 4-week-old Crouzon mice, and white bins denote 6-week-old Crouzon mice.

References

[1] 3dMD. http://www.3dmd.com/3dMDface/, 2012.

[2] A.H. Abbott, D.J. Netherway, D.B. Niemann, B. Clark, M. Yamamoto, J. Cole, A. Hanieh, M.H. Moore, and D.J. David. CT-determined intracranial volume for a normal population. *Journal of Craniofacial Surgery*, 11(3):211–23, 2000.

[3] N. Acer, B. Sahin, O. Baş, T. Ertekin, and M. Usanmaz. Comparison of three methods for the estimation of total intracranial volume: Stereologic, planimetric, and anthropometric approaches. *Annals of Plastic Surgery*, 58(1):48–53, 2007.

[4] Pancreatic Cancer Action. http://pancreaticcanceraction.org/facts-figures/how_pancreatic_cancer_is_diagnosed/mri-scan/, 2012.

[5] D. Adams, F. J. Rohlf, and D. Slice. Geometric morphometrics: Ten years of progress following the 'revolution'. *Italian Journal of Zoology*, 71(1):5–16, 2004.

[6] A.L. Albright and R.P. Byrd. Suture pathology in craniosynostosis. *Journal of Neurosurgery*, 54:384–7, 1981.

[7] P. J. Anderson, D. J. Netherway, A. Abbott, and D. J. David. Intracranial volume measurement of metopic craniosynostosis. *J. Craniofacial Surg.*, 15(6):1014–1016, 2004.

[8] P. J. Anderson, D. J. Netherway, K. McGlaughlin, and D. J. David. Intracranial volume measurement of sagittal craniosynostosis. *Journal of Clinical Neuroscience*, 14(5):455–458, 2007.

[9] P.R. Andresen, F.L. Bookstein, K. Conradsen, B.K. Ersbøll, J.L. Marsh, and S. Kreiborg. Surface-bounded growth modeling applied to human mandibles. *IEEE Transactions on Medical Imaging*, 19(11):1053–1063, 2000.

[10] J. Ashburner and K.J. Friston. Voxel-based morphometry-the methods. *NeuroImage*, 11(6):805 – 821, 2000.

[11] J. Ashburner and K.J. Friston. Unified segmentation. *NeuroImage*, 26(3):839–851, 2005.

[12] C. Asher-McDade, V. Brattström, E. Dahl, J. McWilliam, K. Mølsted, D. A. Plint, B. Prahl-Andersen, G. Semb, W. C. Shaw, and R. P. S. The. A six-centre international study of treatment outcome in patients with clefts of the lip and palate: part 4: assessment of nasolabial appearance. *Cleft Palate-Craniofacial Journal*, 29:409–412, 1992.

[13] A. E. Athanasiou, editor. *Orthodontic Cephalometry*. Mosby-Wolfe, 1995.

[14] A.S. Aylsworth. Genetic considerations in craniofacial birth defects. In T.A. Turvey, K.W.L. Vig, and R.J. Fonseca, editors, *Facial clefts and craniosynostosis. Principles and Management*, pages 76–94. W.B. Saunders Company, Philadelphia, 1996.

[15] D.J. Baxter and M.M. Shroff. Developmental maxillofacial anomalies. *Semin Ultrasound CT MR*, 32(6):555 – 568, 2011.

[16] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(5):289–300, 1995.

[17] S. Berkowitz. The effect of clefting of the lip and palate on the palatal arch form in cleft lip and palate. In S. Berkowitz, editor, *Diagnosis and management of cleft lip and palate*, pages 43–54. Springer, Berlin, 2006.

[18] M. Bern and D. Eppstein. Mesh generation and optimal triangulation. In D.-Z. Du and F. Hwang, editors, *Computing in Euclidean Geometry*. World Scientific, Lecture Notes Series on Computing, vol. 1 edition, 1992.

[19] J. Besag. Towards bayesian image analysis. *Journal of Applied Statistics*, 16(3):395–407, 1989.

[20] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. *Proc. Int'l Conf. Computer Vision*, pages 489–495, 1999.

[21] C.M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[22] J.M. Bland and D.G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 1:307–310, 1986.

[23] F.L. Bookstein. *Morphometric Tools for Landmark data: Geometry and Biology*. Cambridge University Press, Cambridge, USA, 1991.

[24] Y. Boykov and M.-P. Jolly. Interactive organ segmentation using graph cuts. *Proc. Medical Image Computing and Computer-Assisted Intervention*, pages 276–286, 2000.

[25] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. *Proc. Int'l Conf. Computer Vision*, pages 105–112, 2001.

[26] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[27] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 648–655, 1998.

[28] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Proc. IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[29] R. Brandt, T. Rohlfing, J. Rybak, S. Krofczik, A. Maye, M. Westerhoff H.-C. Hege, and R. Menzel. Three-dimensional average-shape atlas of the honeybee brain and its applications. *The Journal of Comparative Neurology*, 492:1–19, 2005.

[30] J.A. Bærentzen, J. Graversen, F. Anton, and H. Aanæs. *Guide to computational geometry processing*. Springer, 2012.

[31] B. H. Broadbent. A new X-ray technique and its application to orthodontia. *Angle Orthod*, 1:45–66, 1931.

[32] I. Bugaighis, P. O'Higgins, B. Tiddeman, C. Mattick, O. Ben Ali, and R. Hobson. Three-dimensional geometric morphometrics applied to the study of children with cleft lip and/or palate from the north east of england. *Eur J Orthod*, 32(5):514–521, 2010.

[33] M.J. Burke and S. Williams K.R. Winston. Normal suture fusion and the etiology of single sutural craniosynostosis: the microspicule hypothesis. *Pediatric Neurosurgery*, 22:241–6, 1995.

[34] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M.B. Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Comput Methods Programs Biomed*, 104(3):e158–77, 2011.

[35] J.M. Carstensen. *Description and simulation of visual texture.* PhD thesis, IMSOR, Technical University of Denmark, 1992.

[36] L.H.S. Cevidanes, L'T. J. Bailey, S.F. Tucker, M.A. Styner, A. Mol, C.L. Phillips, W.R. Proffit, and T.Turvey. Three-dimensional cone-beam computed tomography for assessment of mandibular changes after orthognathic surgery. *American Journal of Orthodontics and Dentofacial Orthopedics*, 131(1):44 – 50, 2007.

[37] H.S. Choi, H.G. Choi, S.H. Kim, H.J. Park, D.H. Shin, D.I. Jo, C.K. Kim, and K.I. Uhm. Influence of the alveolar cleft type on preoperative estimation using 3d ct assessment for alveolar cleft. *Arch Plast Surg*, 39(5):477–82, 2012.

[38] G.E. Christensen, A.A. Kane, J.L. Marsh, and M.W. Vannier. Synthesis of an individualized cranial atlas with dysmorphic shape. In *MMBIA '96: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, page 309, Washington, DC, USA, 1996. IEEE Computer Society.

[39] M.M. Cohen and R.E. MacLean, editors. *Craniosynostosis. Diagnosis, Evaluation and Management.* Oxford University Press, second edition, 2000.

[40] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality image registration using information theory. *Proceedings of Information Processing in Medical Imaging*, pages 263–274, 1995.

[41] M.J. Cook. *The anatomy of the laboratory mouse*, pages 11–15. Academic Press, 1965.

[42] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[43] O. Crouzon. Dysostose cranio-faciale héréditère. *Bull Mem Soc Méd Hôp Paris*, 33:545–555, 1912.

[44] M. Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J.G. Villemure, and J.P. Thiran. Atlas-based segmentation of pathological mr brain images using a model of lesion growth. *IEEE Transactions on Medical Imaging*, 23:1301–1314, 2004.

[45] E. Dahl and S. Kreiborg. Craniofacial malformations. In B. Thilander and O. Rönning, editors, *Introduction to Othodontics*, chapter XI, pages 239–246. Förlagshuset Gothia AB with LIC Förlag, Stockholm, 2 edition, 1995.

[46] G. Dahlberg. *Statistical Methods for Medical and Biological Students*. London: Georg Allen and Unwin Ltd, 1940.

[47] T.A. Darvann. *Methods for measurement and analysis of craniofacial morphology and growth in children with cleft lip and palate*. PhD thesis, Technical University of Denmark, Lyngby, Denmark, 2003.

[48] T.A. Darvann. Landmarker: A vtk-based tool for landmarking of polygonal surfaces. In K. Takada and S. Kreiborg, editors, *Silico Dentistry - The Evolution of Computational Oral Health Science*, pages 160–62. Medigit, Osaka, Japan, 2008.

[49] T.A. Darvann, N.V. Hermann, S. Demant, P. Larsen, H. Olafsdottir, S.S. Thorup, M. Zak, A.B. Lipira, A.A. Kane, D. Govier, H. Schatz, D. Rueckert, and S. Kreiborg. Automated quantification and analysis of facial asymmetry in children with arthritis in the temporomandibular joint. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1193–1196, 2011.

[50] A.S. Dekaban. Table of cranial and orbital measurements, cranial volume and derived indexes in male and females from 7 days to 20 years of age. *Annals of Neurology*, 2(6):485–491, 1977.

[51] Dentronic. http://www.dentronic.dk/PM_Promax.aspx?ID=631, 2012.

[52] L.R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[53] I.L. Dryden and K.V. Mardia. *Statistical Shape Analysis*. JohnWiley & Sons, Chichester, England, UK, 1998.

[54] V. Duay, N. Houhou, and J.P. Thiran. Atlas-based segmentation of medical images locally constrained by level sets. *International Conference on Image Processing 2005 Proceedings*, 2:1286–1289, 2006.

[55] N. Duchateau, M. De Craene, G. Piella, E. Silva, A. Doltra, M. Sitges, B.H. Bijnens, and A.F. Frangi. A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. *Med Image Anal*, 15(3):316–328, 2011.

[56] J. Duchon. Interpolation des function de deux variables suivant les principes de la flexion des plaques minces. *RAIRO Analyse Numérique*, 10(3):5–12, 1976.

[57] B. Efron. Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*, 99(465), 2004.

[58] J Eritaia, S.J. Wood, G.W. Stuart, N. Bridle, P. Dudgeon, P. Maruff, D. Velakoulis, and C. Pantelis. An optimized method for estimating intracranial volume from magnetic resonance images. *Magn Reson Med.*, 44(6):973–977, 2000.

[59] K. Erleben, J. Sporring, K. Henriksen, and H. Dohlmann. *Physics-based animation*, pages 288–292. Charles River Media, Inc., 2005.

[60] B.K. Ersbøll and K. Conradsen. *An introduction to statistics, vol. 2*, pages 51–68. IMM, Technical University of Denmark, seventh edition, 2007.

[61] S.F. Eskildsen and L.R. Østergaard. Active surface approach for extraction of the human cerebral cortex from mri. In *Proceedings of the 9th international conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part II*, MICCAI'06, pages 823–830, Berlin, Heidelberg, 2006. Springer-Verlag.

[62] V.P. Eswarakumar, M.C. Horowitz, R. Locklin, G.M. Morriss-Kay, and P. Lonai. A gain-of-function mutation of Fgfr2c demonstrates the roles of this receptor variant in osteogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34):12555–12560, 2004.

[63] M. Feichtinger, R. Mossböck, and H. Kärcher. Assessment of bone resorption after secondary alveolar bone grafting using three-dimensional computed tomography: a three-year study. *Cleft Palate Craniofac J*, 44(2):142–148, 2007.

[64] L.A. Feldkamp, L.C. Davis, and J.W. Kress. Practical cone-beam algorithm. *JOSA A*, 1(6):612–619, 1984.

[65] D.R. Forsey and R.H. Bartels. Hierarchical B-spline refinement. *ACM Transactions on Computer Graphics*, 22(4):205 – 212, 1988.

[66] H. Förstl, R Zerfab, C Geiger-Kabisch, H Sattel, C Besthorn, and F Hentschel. Brain atrophy in normal ageing and alzheimer's disease. volumetric discrimination and clinical correlations. *The British journal of psychiatry*, 167(739–746):6, 1995.

[67] P. Frederiksen. Image registration and warping. In J.M. Carstensen, editor, *Image analysis, vision and computer graphics*, pages 203–212. Informatics and Mathematical Modelling, Technical University of Denmark, second edition, 2002.

[68] G. Frongia, M. G. Piancino, and P. Bracco. Cone-beam computed tomography: Accuracy of three-dimensional cephalometry analysis and influence of patient scanning position. *Journal of Craniofacial Surgery*, 23:1038–1043, 2012.

[69] D. Gault, F. Brunelle, D. Renier, and D. Marchac. The calculation of intracranial volume using CT scans. *Childs Nerv Syst*, 4(5):271–273, 1988.

[70] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[71] A.K. Gosain, J.G. McCarthy, P. Glatt, D. Staffenberg, and R.G. Hoffmann. A study of intracranial volume in Apert syndrome. *Plast Reconstr Surg*, 95(2):284–295, 1995.

[72] P.J. Green and B.W. Silverman. *Nonparametric regression and generalized linera models. A roughness penalty approach.* Chapman and Hall/CRC, 1994.

[73] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *J. Royal Statistical Soc., Series B*, 51(2):271–279, 1989.

[74] A. Guimond, J. Meunier, and J.-P. Thirion. Average brain models: A convergence study. *Computer vision and image understanding*, 77(2):192–210, 2000.

[75] J.V. Hajnal, D.L.G. Hill, and D.J. Hawkes, editors. *Medical Image Registration*, pages 11–71, 281–303. CRC Press, 2001.

[76] A. Hald. *Statistical Theory with Engineering Applications*. John Wiley and Sons Inc., New York and London, 1952.

[77] Y. Hamada, T. Kondoh, K. Noguchi, M. Iino, H. Isono, H. Ishii, A. Mishima, K. Kobayashi, and K. Seto. Application of limited cone beam computed tomography to clinical assessment of alveolar bone grafting: a preliminary report. *Cleft Palate Craniofac J*, 42(2):128–137, 2005.

[78] M.S. Hansen, S.S. Thorup, and S.K. Warfield. Polyaffine parametrization of image registration based on geodesic flows. In *Mathematical Methods in Biomedical Image Analysis (MMBIA), 2012 IEEE Workshop on*, pages 289–295, 2012.

[79] L.G. Hanson. Introduction to magnetic resonance imaging techniques, 2009. http://eprints.drcmr.dk/37/.

[80] N.V. Hermann. *Craniofacial Morphology and Growth in Infants and Young Children with Cleft Lip And Palate*. PhD thesis, Department of Pediatric dentistry, School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark, 2000.

[81] N.V. Hermann, B.L. Jensen, E. Dahl, T.A. Darvann, and S. Kreiborg. A method for three-projection infant cephalometry. *Cleft Palate Craniofacial Journal*, 38(4):299–316, 2001.

[82] K.B. Hilger, R. Larsen, and M.C. Wrobel. Growth modeling of human mandibles using non-euclidean metrics. *Medical Image Analysis*, 7:425–433, 2003.

[83] H. Hofrath. Die bedeutung der röntgenfern- und abstandsaufname für die diagnostik der kieferanomalien. *Fortschr. Othod.*, 1:232–258, 1931.

[84] A. Hrdlicka. *Anthropometry*. The Wistar Institute of Anatomy and Biology, Philadelphia, 1920.

[85] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the Hausdorff distance. *PAMI, IEEE Trans. on*, 15(9):850 –863, 1993.

[86] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. *Proc. European Conf. Computer Vision*, pages 232–248, 1998.

[87] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 125–131, 1998.

[88] R. Jenkins, N.C. Fox, A.M. Rossor, R.J. Harvey, and M.N. Rossor. Intracranial volume and alzheimer's disease: Evidence against the cerebral reserve hypothesis. *Arch. Neurol.*, 57:220–224, 2000.

[89] B.L. Jensen, S. Kreiborg, E. Dahl, and P. Fogh-Andersen. Cleft lip and palate in Denmark, 1976-1981: epidemiology, variability, and early somatic development. *Cleft Palate Journal*, 25:258–269, 1988.

[90] S. Joshi, B. Davis, M. Jomierc, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.

[91] M. M. Cohen Jr. and S. Kreiborg. Birth prevalence studies of crouzon syndrome: comparison of direct and indirect methods. *Clinical Genetics*, 41:12–15, 1992.

[92] S.S. Keller and N. Roberts. Measurement of brain volume using mri: software, techniques, choices and prerequisites. *J Anthropol Sci*, 87:127 – 151, 2009.

[93] Sick Kids. http://www.sickkids.ca/Craniofacial/What-we-do/Craniofacial-Conditions/Craniosynostosis/Unicoronal-Synostosis/index.html, 2012.

[94] J. Kim, J. Fisher, A. Tsai, C. Wible, A. Willsky, and W. Wells. Incorporating spatial priors into an information theoretic approach for FMRI data analysis. *Proc. Medical Image Computing and Computer-Assisted Intervention*, pages 62–71, 2000.

[95] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. *Proc. Int'l Conf. Computer Vision*, pages 1033–1040, 2003.

[96] J. Kim and R. Zabih. Automatic segmentation of contrast-enhanced image sequences. *Proc. Int'l Conf. Computer Vision*, pages 502–509, 2003.

[97] A. Klein, J. Andersson, B.A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G.E. Christensen, D.L. Collins, J. Gee, P. Hellier, J.H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R.P. Woods, J.J. Mann, and R.V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802, 2009.

[98] B. Knutson, R. Momenan, R.R. Rawlings, G. W. Fong, and D Hommer. Negative association of neuroticism with brain volume ratio in healthy humans. *Biol. Psychiatry*, 50(9):685–690, 2001.

[99] P. Kohli and P.H.S. Torr. Dynamic graph cuts for efficient inference in markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2079–2088, 2007.

[100] V. Kolmogorov and R. Zabih. Visual correspondence with occlusions using graph cuts. *Proc. Int'l Conf. Computer Vision*, pages 508–515, 2001.

[101] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *PAMI, IEEE Trans. on*, 26(2):147–159, 2004.

[102] S. Kreiborg. *Crouzon syndrome. A clinical and roentgencephalometric study.* Odontologisk boghandels forlag, 1981.

[103] S. Kreiborg. Postnatal growth and development of the craniofacial complex in premature craniosynostosis. In M.M. Cohen and R.E. MacLean, editors, *Craniosynostosis. Diagnosis, Evaluation, and Management*, pages 158–74. Oxford University Press, 2 edition, 2000.

[104] S. Kreiborg, E. Dahl, and U. Prydsø. A unit for infant roentgencephalometry. *Dentomaxillofac Radiol*, 6:107–111, 1977.

[105] H. Ólafsdóttir. *Analysis of Craniofacial Images using Computational Atlases and Deformation Fields*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2007. IMM-PHD-2007-187.

[106] Y.F. Lam, D.F. Gillies, D. Rueckert, P. Charters, P. Groom, and S. Roughley. A generic anatomical model of the human mandible. In M.H. Hamza, editor, *Proceedings of the IASTED International Conference on Biomechanics*, pages 247–251. ACTA Press, 2003.

[107] R. Larsen. 02505 course note - medical image analysis. rl(at)imm.dtu.dk, 2008.

[108] A. Lee and K. Pearson. Data for the problem of evolution in man-a first study of the correlation of the human skull. *Philosophical Transactions of Royal Society*, 196a:225–264, 1901.

[109] S. Lele and J. Richtsmeier. Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology*, 86:415–427, 1991.

[110] R. Leonardi, D. Giordano, F. Maiorana, and C. Spampinato. Automatic cephalometric analysis. *Angle Orthodontist*, 78(1):145–150, 2008.

[111] H. Li, Y. Yang, Y. Chen, Y. Wu, Y. Zhang, D. Wu, and Y. Liang. Three-dimensional reconstruction of maxillae using spiral computed tomography and its application in postoperative adult patients with unilateral complete cleft lip and palate. *J Oral Maxillofac Surg*, 69(12):e549–557, 2011.

[112] K. Li, X. Wu, D.Z. Chen, and M. Sonka. Optimal surface segmentation in volumetric images-a graph-theoretic approach. *PAMI, IEEE Trans. on*, 28(1):119–134, 2006.

[113] Wikipedia Cleft lip and palate. http://en.wikipedia.org/wiki/Cleft_lip_and_palate, 2012.

[114] W.E. Lorensen and H.E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, 21(4):163–169, 1987.

[115] M. Lorenzo-Valdes, D. Rueckert, and R. Mohiaddin G. Sanchez-Ortiz. Segmentation of cardiac mr images using the em algorithm with a 4d probabilistic atlas and a global connectivity filter. *Engineering in Medicine and Biology Society, 2003 Proceedings of the 25th Annual International Conference of the IEEE*, 1:626–629, 2003.

[116] J. E. Losee and R. E. Kirschner, editors. *Comprehensive Cleft Care*. McGraw-Hill Medical, New York, Chicago, San Francisco, Lisbon, London, Madrid, Mexico City, Milan, New Delhi, San Juan, Seoul, Singapore, Sydney, Toronto, 2008.

[117] PD Lyden, R Zweifler, Z Mahdavi, and L. Lonzo. A rapid, reliable, and valid method for measuring infarct and brain compartment volumes from computed tomographic scans. *Stroke*, 25(12):2421–2428, 1994.

[118] W.R. Macdonell. A study of the variation and correlation of the human skull with special reference to english crania. *Biometrika*, 3:191–244, 1904.

[119] F. Maes, A. Collignon, D. Vandermeulen, G. Marechal, and R. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.*, 16(2):187–198, 1997.

[120] K. Y. Manjunath. Estimation of cranial volume - an overview of methodologies. *Journal of the Anatomical Society of India*, 51(1):85–91, 2002.

[121] M. Mazonakis, S. Karampekios, J. Damilakis, A. Voloudaki, and N. Gourtsoyiannis. Stereological estimation of total intracranial volume on CT. *Eur. Radiol.*, 14:1285–1290, 2004.

[122] G.R. McGhee. *Theoretical Morphology: the Concept and its Applications*. Columbia University Press, 1999.

[123] T. McInerney and D. Terzopoulos. T-snakes: topology adaptive snakes. *Med Image Anal*, 4(2):73–91, 2000.

[124] J. Meinguet. Multivariate interpolation at arbitrary points made simple. *Journal of Applied Mathematics and Physics (ZAMP)*, 30:292–304, 1979.

[125] M.K. Misztal. *Deformable Simplicial Complexes*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2010.

[126] M.K. Misztal and J.A. Bærentzen. Topology adaptive interface tracking using the deformable simplicial complex. *ACM Transactions on Graphics*, 31(3):Article 24, 2012.

[127] M.K. Misztal, R. Bridson, K. Erleben, J.A. Bærentzen, and F. Anton. Optimization-based fluid simulation on unstructured meshes. In *7th Workshop on Virtual Reality Interaction and Physical Simulation, VRIPHYS*. 2010.

[128] G. M. Morriss-Kay and A.O.M. Wilkie. Growth of the normal skull vault and its alteration in craniosynostosis: insights from human genetics and experimental studies. *Journal of Anatomy*, 207:637–653, 2005.

[129] J.C. Murray. Gene/environment causes of cleft lip and/or palate. *Clinical Genetics*, 61(4):248–256, 2002.

[130] Medical School Lecture Notes. http://imueos.wordpress.com/2010/10/10/base-of-the-skull/, 2013.

[131] The Encyclopedia of Science. http://www.daviddarling.info/encyclopedia/F/fontanelle_skull.html, 2012.

[132] H. Ólafsdóttir, T.A. Darvann, N.V. Hermann, P. Larsen, D. Govier, A.A. Kane, H.D. Hove, J. Hukki, R. Larsen, B.K. Ersbøll, and S. Kreiborg. Non-rigid registration in craniofacial image analysis: Application to the study of unicoronal synostosis. In K. Takada and S. Kreiborg, editors, *Silico Dentistry - The Evolution of Computational Oral Health Science*, pages 98–101. Osaka: Medigit, 2008.

[133] H. Ólafsdóttir, T.A. Darvann, N.V. Hermann, E. Oubel, B.K. Ersbøll, A.F. Frangi, P. Larsen, C.A. Perlyn, G.M. Morriss-Kay, and S. Kreiborg. Computational mouse atlases and their application to automatic assessment of craniofacial dysmorphology caused by the crouzon mutation Fgfr2 C342Y. *Journal of Anatomy*, 211(1):37–52, 2007.

[134] C.G. Ortiz and A.L. Martel. Automatic atlas-based segmentation of the breast in MRI for 3D breast volume computation. *Med. Phys*, 39(10):5835–5848, 2012.

[135] S.J. Osher and R.P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2002.

[136] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Systems man and cybernetics*, 9(1):62–66, 1979.

[137] H. Park, P. Bland, and C. Meyer. Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Transactions on Medical Imaging*, 22:483–492, 2003.

[138] R.R. Paulsen and T.B. Moeslund. *Introduction to Medical Image Analysis*. DTU Informatics, 2012.

[139] G. Pengas, J. M. S. Pereira, G. B. Williams, and P. J. Nestor. Comparative reliability of total intracranial volume estimation methods and the influence of atrophy in a longitudinal semantic dementia cohort. *Journal of Neuroimaging*, 19(1):37–46, 2009.

[140] C.A. Perlyn, V.B. DeLeon, C. Babbs, D. Govier, L. Burell, T.A. Darvann, S. Kreiborg, and G.M. Morriss-Kay. The craniofacial phenotype of the crouzon mouse: Analysis of a model for syndromic craniosynostosis using 3D micro CT. *Cleft Palate-Craniofacial Journal*, 43(6):740–747, 2006.

[141] C.A. Perlyn, G.M. Morriss-Kay, T.A. Darvann, M. Tenenbaum, and D.M. Ornitz. A model for the pharmacological treatment of crouzon syndrome. *Neurosurgery*, 59(1):210–215, 2006.

[142] D. Perperidis, R. Mohiaddin, and D. Rueckert. Construction of a 4d statistical atlas of the cardiac anatomy and its use in classification. *Eighth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2005), LNCS*, 3750:402–410, 2005.

[143] J.M. Plooij, T.J.J. Maal, P. Haers, W.A. Borstlap, A.M. Kuijpers-Jagtman, and S.J. Bergé. Digital three-dimensional image fusion processes for planning and evaluating orthodontics and orthognathic surgery. a systematic review. *Int J Oral Maxillofac Surg.*, 40(4):341–52, 2011.

[144] H. Poll. Ein neuer apparat zur bestimmung der schädelkapazitat. *Ver. Berl. Ges. Anthropol.*, pages 615–620, 1896.

[145] J.C. Posnick, U. Bite, P. Nakano, J. Davis, and D. Armstrong. Indirect intracranial volume measurements using ct scans: clinical applications for craniosysnostosis. *Plast Reconstr Surg*, 89(1):34–45, 1992.

[146] J.L. Prince and J.M. Links, editors. *Medical Imaging - Signals and Systems*. Pearson Prentice Hall Bioengineering, 2006.

[147] C.R. Rao and S. Suryawanshi. Statistical analysis of shape through triangulation of landmarks: a study of sexual dimorphism in hominids. *Proc. Natl. Acad. Sci. U.S.A.*, 95:4121–4125, 1998.

[148] W. Reardon, R.M. Winther, P. Rutland, L.J. Pulleyn, B.M. Jones, and S. Malcolm. Mutations in the fibroblast growth factor receptor type 2 gene cause crouzon syndrome. *Nature Genetics*, 8:98–103, 1994.

[149] R. Recinos, C. Hanger, R. Schaefer, C. Dawson, and A. Gosain. Microfocal CT: A method for evaluating murine cranial sutures in situ. *Journal of Surgical Research*, 116:322–329, 2004.

[150] D.P. Rice, editor. *Craniofacial sutures. Development, disease and treatment*, pages 1–21, 91–106. Karger, 2008.

[151] J. Richtsmeier. Comparative study of normal, crouzon, and apert craniofacial morphology using finite element scaling analysis. *Am J Phys Anthropol*, 74:473–493, 1987.

[152] J.T. Richtsmeier, L.L. Baxter, and R.H. Reeves. Parallels of craniofacial maldevelopment in down syndrome and Ts65Dn mice. *Development Dynamics*, 217:137–145, 2000.

[153] F.J. Rohlf. Fitting curves to outlines. In F.J. Rohlf and F.L. Bookstein, editors, *Proceedings of the Michigan morphometrics workshop. Spec. Publ. No. 2.*, pages 167–177. University of Michigan Museum of Zoology, Ann Arbor, 1990.

[154] F.J. Rohlf. Shape statistics: Procrustes superimpositions and tangent spaces. *J. Class.*, 16:197–223, 1999.

[155] F.J. Rohlf. On the use of shape spaces to compare morphometric methods. *Hystrix*, 11:9–25, 2000.

[156] F.J. Rohlf. Bias and error in estimates of mean shape in morphometrics. *J. Human Evol.*, 44:665–683, 2003.

[157] F.J. Rohlf and L.F. Marcus. A revolution in morphometrics. *Trends Ecol. Evol.*, 8:129–132, 1993.

[158] T. Rohlfing. Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. *IEEE transactions on medical imaging*, 31(2):153–163, 2012.

[159] S. Roy. Stereo without epipolar lines: A maximum flow formulation. *Int'l J. Computer Vision*, 1(2):1–15, 1999.

[160] S. Roy and I. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. *Proc. Int'l Conf. Computer Vision*, pages 492–499, 1998.

[161] D. Rueckert, A.F. Frangi, and J.A. Schnabel. Automatic construction of 3-D statistical deformation models of the brain using non-rigid registration. *IEEE transactions on Medical Imaging*, 22(8):1014–1025, august 2003.

[162] D. Rueckert and J.A. Schnabel. Image Registration Toolkit (IRTK). http://www.doc.ic.ac.uk/~dr/software/index.html, February 2008.

[163] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[164] Uspenskii. S. A new method for measuring cranial capacity. *American Journal of Physical Anthropology*, 22:115–117, 1954.

[165] H. M. Saal. Classification and description of nonsyndromic clefts. In D. F. Wyszynski, editor, *Cleft lip and palate from origin to treatment*, chapter V, pages 47–52. Oxford University Press, 2002.

[166] S.A. Sadananthan, W. Zheng, M.W. Chee, and V. Zagorodnov. Skull stripping using graph cuts. *Neuroimage*, 49(1):225 – 239, 2010.

[167] C. Sadowsky, H. Aduss, and S. Pruzansky. The soft tissue profile in unilateral clefts. *Angle Orthodontist*, 43:233–246, 1973.

[168] B. Sahin, N. Acer, O.F. Sonmez, M. Emirzeoglu, H. Basaloglu, A. Uzun, and S. Bilgic. Comparison of four methods for the estimation of intracranial volume: A gold standard study. *Clinical Anatomy*, 20:766–773, 2007.

[169] J. Schäfer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1):32, 2005.

[170] J.A. Schnabel, D. Rueckert, M. Quist, J.M. Blackall, A.D. Castellano-Smith, T. Hartkens, G.P. Penney, W.A. Hall, H. Liu, C.L. Truwit, F.A. Gerritsen, D.L.G. Hill, and D.J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In W.J. Niessen and M.A. Viergever, editors, *MICCAI*, volume 2208 of *Lecture Notes in Computer Science*, pages 573–581. Springer, 2001.

[171] P.W. Schofield, R.E. Mosesson, Y. Stern, and R. Mayeux. The age at onset of Alzheimer's disease and an intracranial area measurement: A relationship. *Archives of Neurology*, 52(1):95–98, 1995.

[172] W. Schroeder, K. Martin, and B. Lorensen. *The Visualization Toolkit. An object-oriented approach to 3D graphics*. Kitware, 3 edition, 2002.

[173] K. Schwenzer-Zimmerer, D. Chaitidis, I. Boerner, L. Kovacs, N.F. Schwenzer, C. Holberg, and H.F. Zeilhofer. Systematic contact-free 3D topometry of the soft tissue profile in cleft lips. *Cleft Palate Craniofac J*, 45(6):607 – 613, 2008.

[174] Sedentexct. http://www.sedentexct.eu/content/technical-description-cbct, 2012.

[175] S. Sgouros, A. D. Hockley, J. H. Goldin, M. J. C. Wake, and K. Natarajan. Intracranial volume change on childhood. *Journal of Neurosurgery*, 91:610–616, 1999.

[176] S. Sgouros, A. D. Hockley, J. H. Goldin, M. J. C. Wake, and K. Natarajan. Intracranial volume change on craniosynostosis. *Journal of Neurosurgery*, 91:617–625, 1999.

[177] R. Shekhar and V. Zagrodsky. Mutual information-based rigid and nonrigid registration of ultrasound volumes. *IEEE Transactions on Medical Imaging*, 21:9–22, 2002.

[178] J.R. Shewchuk. Tetrahedral mesh generation by delaunay refinement. In *Proceedings of the 14th annual symposium on Computational geometry*, pages 86–95. ACM, 1998.

[179] Z. Smahel. Variations in craniofacial morphology with severity of isolated cleft palate. *Cleft Palate-Craniofacial Journal*, 21:140–158, 1984.

[180] S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, 2002.

[181] B. Solow. *The pattern of craniofacial associations: A morphological and methodological correlation and factor analysis study on young male adults*, pages 1–174. Acta Odontol Scand, 24 (Suppl 46), 1966.

[182] B. Solow and S. Kreiborg. A cephalometric unit for research and hospital environments. *Eur J Orthod*, 10:346–352, 1988.

[183] M. Sonka and J.M. Fitzpatrick, editors. *Handbook of medical imaging. Medical image processing and analysis*, volume 2. SPIE Press, 2000.

[184] M. Sonka and J.M. Fitzpatrick, editors. *Medical Imaging - Vol. 2. Medical Image Processing and Analysis*. SPIE - The International Society for Optical Engineering, 2000.

[185] J. Staubesand, editor. *Sobotta. Atlas of the human anatomy.*, volume 1. Urban and Schwarzenberg, 11 edition, 1989.

[186] T.D. Stewart. Cranial capacity studies. *American journal of Physical Anthropology*, XVIII(3):337–361, 1934.

[187] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[188] StudyBlue. Lecture 16 3/29/2011. http://www.studyblue.com/notes/note/n/lecture-16-3292011/deck/868786, 2012.

[189] G. Subsol, J.P. Thirion, and N. Ayache. A scheme for automatically building three-dimensional morphometric anatomical atlases: application to a skull atlas. *Medical Image Analysis*, 2(1):37–60, 1998.

[190] The Jackson Laboratory The Craniofacial Home. http://craniofacial.jax.org/characteristics.html, 2012.

[191] D.W. Thompson. *On Growth and Form*. Dover reprint of 1942 2nd ed., 1992.

[192] S.S. Thorup. Quantification of craniofacial growth in mice with craniofacial dysmorphology caused by the Crouzon mutation Fgfr2 C342Y. Master's thesis, Technical University of Denmark, Lyngby, Denmark, 2008.

[193] S.S. Thorup, T.A. Darvann, H. Ólafsdóttir, N.V. Hermann, R.R. Paulsen, P. Larsen, C.A. Perlyn, R. Larsen, and S. Kreiborg. Quantification of craniofacial growth in mice with craniofacial dysmorphology caused by the crouzon mutation fgfr2^{C342Y}. *Submitted to J. Anatomy*, 2013.

[194] S.S. Thorup, R. Larsen, T.A. Darvann, H. Ólafsdóttir, R.R. Paulsen, N.V. Hermann, P. Larsen, C.A. Perlyn, and S. Kreiborg. Automatic assessment of craniofacial growth in a mouse model of crouzon syndrome. In *American Cleft Palate-Craniofacial Association: 66th Annual Meeting*, pages 74–75. 2009.

[195] S.S. Thorup, H. Ólafsdóttir, T.A. Darvann, N.V. Hermann, P. Larsen, R.R. Paulsen, C.A. Perlyn, G.M. Morriss-Kay, S. Kreiborg, and R. Larsen. Multivariate analysis of variance: Finding significant growth in mice with craniofacial dysmorphology caused by the crouzon mutation. In B.K. Ersbøll; G. Guillot, editor, *The Eighth French-Danish Workshop on Spatial Statistics and Image Analysis in Biology: Book of Abstracts*. 2010.

[196] T. Thyagarajan, S. Totey, M. Danton, and A. Kulkarni. Genetically altered mouse models: the good, the bad, and the ugly. *Critical reviews in oral biology and medicine*, 2003.

[197] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *Med. Imag., IEEE Trans. on*, 18(10):897–908, 1999.

[198] O.J. van Vlijmen, M.A. Kuijpers, S.J. Bergé, J.G. Schols, T.J. Maal, H. Breuning, and A.M. Kuijpers-Jagtman. Evidence supporting the use of cone-beam computed tomography in orthodontics. *J Am Dent Assoc*, 143(3):241–252, 2012.

[199] F. Vegter, J. W. Mulder, and J. J. Hage. Major residual deformities in cleft patients: a new anthropometric approach. *Cleft Palate-Craniofacial Journal*, 34:106–110, 1997.

[200] P.A. Viola and W.M. Wells III. Alignment by maximisation of mutual information. In *Proc. 5th Int. Conf. on Computer Vision*, pages 16–23, 1995.

[201] G. Wahba. *Spline models for observational data*. SIAM, 1990.

[202] J.A. Walker. The ability of geometric morphometric methods to estimate a known covariance matrix. *Syst. Biol.*, 49:686–696, 2000.

[203] H. Wanifuchi, T. Shimizu, and T. Maruyama. Age-related changes in the proportion of intracranial cerebrospinal fluid space measured using volumetric computerized tomography scanning. *J Neurosurg*, 97(3):607–610, 2002.

[204] A. Watson. Embryology, aetiology and incidence. In A. Watson, D. Sell, and P. Grunwell, editors, *Management of cleft lip and palate*, pages 3–15. Whurr, London, 2004.

[205] J. L. Whitwell, W. R. Crum, H. C. Watt, and N. C. Fox. Normalization of cerebral volumes by use of intracranial volume: Implications for longitudinal quantitative MR imaging. *American Journal of Neuroradiology*, 22:1483 – 1489, 2001.

[206] H. Wolf, F Kruggel, A Hensel, L.O. Wahlund, T. Arendt, and H.J. Gertz. The relationship between head size and intracranial volume in elderly subjects. *Brain Res.*, 973:74–80, 2003.

[207] R. Wörtche, S. Hassfeld, C.J. Lux, E. Müssig, F.W. Hensley, R. Krempien, and C. Hofele. Clinical application of cone beam digital volume tomography in children with cleft lip and palate. *Dentomaxillofac Radiol*, 35(2):88–94, 2006.

[208] T. Yasuda, K. Harada, S. Yokoi, and J.-I. Toriwaki. An automatic measurement for intracranial volume using 3-d image processings. *Systems and Computers in Japan*, 23(7):89–99, 1992.

[209] N. Zamora, J.-M. Llamas, R. Cibrian, J.-L. Gandia, and V. Parades. A study of the reproducibility of cephalometric landmarks when undertaking a three-dimentional 3D cephalometric analysis. *Med Oral Patol Cir Bucal.*, 17(4):678–88, 2012.

[210] L. Zhang, E. Homan, and J. Reinhardt. Atlas-driven lung lobe segmentation in volumetric x-ray ct images. *IEEE Transactions on Medical Imaging*, 25:1–16, 2006.

[211] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *Med. Imag., IEEE Trans. on*, 1(20):45–57, 2001.